Spring 2011

Multivariate Nonparametric Estimation of Value at Risk and Expected Shortfall for Nonlinear Returns Using Extreme Value Theory

Ryan Brauchler
University of Colorado Boulder

Multivariate Nonparametric Estimation of Value at Risk and Expected Shortfall for Nonlinear Returns Using Extreme Value Theory

By,

Ryan Brauchler

Primary Advisor: Professor Carlos Martins-Filho, Economics Department

Honors Committee Member: Professor Terra McKinnish, Economics Department

Outside Reviewer: Professor Karl Gustafson, Math Department

Undergraduate Departmental Honors Thesis

Department of Economics

University of Colorado at Boulder

November 4, 2011

Multivariate Nonparametric Estimation of Value at Risk and Expected Shortfall for Nonlinear Returns Using Extreme Value Theory*

Abstract

The catastrophic failures of risk management systems in 2008 bring to the forefront the need for accurate and flexible estimators of market risk. Despite advances in the theory and practice of evaluating risk, existing measures are notoriously poor predictors of loss in high-quantile events. To extend the research concerned with modeling extreme value events, we utilize extreme value theory (EVT) to propose a multivariate estimation procedure for value-at-risk (VaR) and expected shortfall (ES) for conditional distributions of a time series of returns on a financial asset. Our approach extends the local linear estimator of conditional mean and volatility used in the conditional heteroskedastic autoregressive nonlinear (CHARN) model proposed by Martins-Filho and Yao (2006) by incorporating an exogenous time series resembling returns on the S&P 500 from January 1950 through September 2011. In combination with EVT, this model estimates the quantiles of the conditional distribution and subsequently the one-day forecasted VaR and ES. We examine the finite sample properties of our method and contrast them with the popular Gaussian GARCH estimator in an extensive Monte Carlo simulation. The method we propose generally outperforms the Gaussian GARCH estimator, particularly in samples greater than 1000. Our results provide evidence of the effect of the curse of dimensionality, which arises because we include a second regressor.

Keywords and phrases. Value at risk, expected shortfall, market risk, nonparametric estimation, extreme value theory, L-moments, nonlinear modeling, Monte Carlo

JEL Classifications. C01, C14, C58, G01, G17, G32.

AMS-MS Classification. 62G05, 62G08, 62G32.

*The author would like to thank Carlos Martins-Filho for his invaluable contributions, suggestions, and teachings during the research process. The author would also like to thank David Keyes, of the University of Colorado Math Department, for his revisions and commentary.

Contents

1 Introduction

"There is always a well-known solution to every human problem - neat, plausible, and wrong."

-H.L. Mencken

After myriad instances of catastrophic failure of risk management systems during the financial crisis of 2008, accurate measurement of the degree to which firms are exposed to market risk became a central concern among internal risk management departments, regulators, and investors. Recent financial reform measures including Basel III, the Volcker Rule, and the sweeping Dodd-Frank Act exemplify the gravity granted to reliably mitigating and accurately measuring risk. Accurate estimation of the market risk to which financial institutions are exposed gives policymakers and portfolio managers insight into capital adequacy requirements which they can use to make better-informed decisions. This paper aims to construct alternative estimators for value-at-risk (VaR) and expected shortfall (ES) to outperform those existing in the literature and provide risk managers and legislators with a better predictor of performance in extreme scenarios, thereby helping forecast, mitigate, and manage risk.

The challenge of synthetically measuring the market risk faced by a firm with a single figure gave rise to VaR (JPMorgan (1996)) and ES (Artzner et al. (1999)). VaR estimates the maximum financial loss on a portfolio over a given time horizon (usually 24 hours) under a specified confidence level (Jorion (2001)). By contrast, ES, also known as Conditional VaR or TailVaR, considers the expected value of all losses exceeding a quantile prescribed by a level of confidence over a specified time interval (Acerbi and Tasche (2002)). Statistically, VaR is a quantile and ES is the expected value of a random variable exceeding a quantile. Since their conception, VaR and ES have been both praised and criticized, and many alternative measures have been proposed in the literature. Though VaR and ES are often adequate risk measures, they are notoriously difficult to estimate for high order quantiles, which is an issue this paper aims to address.

The model proposed in this paper modifies the local linear estimator of conditional mean and volatility used in the conditional heteroskedastic autoregressive nonlinear (CHARN) model proposed by Martins-Filho and Yao (2006) for estimating quantiles of conditional distributions. Hereafter, their original model will be termed the MFY Model. This paper makes two contributions to the literature on VaR and ES estimation. First, we propose the inclusion of an exogenous explanatory variable in the conditional location scale model used in estimation of VaR and ES. Specifically, we consider adding an exogenous series modeled after the returns distribution of the S&P 500 equity index from January 3, 1950 through September 30, 2011. This stochastic variable will act as a control for factors that exhibit significant collinearity with the primary time series. Second, we conduct an extensive Monte Carlo simulation to examine the finite sample properties of our estimator. The Monte Carlo compares the relative performance of our model against the ever popular Gaussian Generalized Autoregressive Conditionally Heteroskedastic (GARCH) model of Bollerslev (1986). Our Monte Carlo study considers several data generating processes (DGPs) that exhibit the empirical properties of financial time series, including "asymmetric conditional volatility, leptokurdicity, infinite past memory and asymmetry of conditional return distributions" (Martins-Filho and Yao (2006)). Performance is measured by root mean squared error (RMSE) and bias.

The remainder of this paper is organized as follows: Section 2 provides a discussion of the statistical methods used in our estimation, EVT, properties of financial assets' returns, and approaches to modeling the returns and volatility of financial assets. Section 3 offers a detailed treatment of the VaR and ES estimation methods we use. Section 4 outlines the design of the Monte Carlo simulation. Section 5 summarizes the results of the Monte Carlo. Section 6 contains a brief conclusion and suggestions for further research.

2 Literature Review

This literature review will focus on the statistical methods used in our estimation procedure for VaR and ES. Since the procedure here modifies the existing MFY procedure and is therefore predetermined, we limit the discussion of previously proposed VaR and ES estimation procedures to ARMA and GARCH variants of first stage estimators. Instead, this section focuses on the mathematical concepts and methods that are essential to understanding the estimation procedure used in our model.

2.1 Data Generating Process

We first define the data generating process used as the basis for the Monte Carlo simulation presented in section 4. It is this process that underlies all the data we use. The DGP that we consider is a modified version of the nonparametric GARCH model proposed by Hafner (1998), studied by Carroll et al. (2002), and utilized by Martins-Filho and Yao (2006). Take $\{Y_t\}$ to be a stochastic process of log-returns on a financial asset where $E(Y_t|Y_{t-1}, D_{t-1}) = 0$ and $E(Y_t^2|Y_{t-1}, D_{t-1}) = \sigma_t^2$ and where D_{t-1} represents lagged returns of an exogenous variable. For our purposes, the exogenous variable mimics the returns distribution of the S&P 500 since January 3, 1950. We assume the returns process evolves as,

$$Y_t = \sigma_t \epsilon_t \text{ for } t = 1, 2, ... \tag{2.1}$$

$$\sigma_t^2 = g(Y_{t-1}, D_{t-1}) + \gamma \sigma_{t-1}^2 \tag{2.2}$$

where $g(x)$ is a positive, twice continuously differentiable function and $0 < \gamma < 1$ is a weighting parameter for the one-period lagged volatility. ϵ_t is a sequence of IID random variables exhibiting a skewed Student-t distribution. ϵ_t is also independent of both Y_{t-1} and D_{t-1}. For the derivation and a discussion of the skewed Student-t density, consult Hansen (1994). The skewed Student-t's PDF, normalized to have $E(\epsilon_t) = 0$ and $Var(\epsilon_t) = 1$, is given by

$$f(x; v, \lambda) = \begin{cases} bc\left(1 + \dfrac{1}{v-2}\left(\dfrac{bx+a}{1+\lambda}\right)^2\right)^{(-v+1)/2} & \text{for } x \geq -a/b \\ bc\left(1 + \dfrac{1}{v-2}\left(\dfrac{bx+a}{1-\lambda}\right)^2\right)^{(-v+1)/2} & \text{for } x \leq -a/b \end{cases} \tag{2.3}$$

where $c \equiv \dfrac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)\sqrt{\pi(v-2)}}$, $a \equiv 4\lambda c \dfrac{v-2}{v-1}$, $b \equiv \sqrt{1 + 3\lambda^2 - a^2}$. The parameter v represents degrees of freedom and λ is the skewness parameter. Note that when $\lambda = 0$, the skewed Student-t becomes a symmetric standardized Student-t.

Patton (2004) derived the VaR (α-quantile) for the skewed Student-t distributed sequence, $\epsilon_{y,t}$, given by

$$q_\epsilon(\alpha) = \begin{cases} \dfrac{1-\lambda}{b}\sqrt{\dfrac{v-2}{v}}F_s^{-1}\left(\dfrac{\alpha}{1-\lambda}, v\right) - \dfrac{a}{b} & \text{for } 0 < \alpha < \dfrac{1-\lambda}{2} \\ \dfrac{1+\lambda}{b}\sqrt{\dfrac{v-2}{v}}F_s^{-1}\left(0.5 + \dfrac{1}{1+\lambda}\left(\alpha - \dfrac{1-\lambda}{2}\right), v\right) - \dfrac{a}{b} & \text{for } \dfrac{1-\lambda}{2} \leq \alpha < 1 \end{cases} \tag{2.4}$$

where F_s^{-1} is the inverse CDF of a random variable with a symmetric Student-t distribution with v degrees of freedom and α confidence level.

Martins-Filho and Yao (2006) derived the Expected Shortfall for the skewed Student-t distributed sequence, ϵ_t, by

$$\begin{aligned} E(\epsilon_t | \epsilon_t > q_\epsilon(\alpha)) = (1 - F(q_\epsilon(\alpha), v))^{-1} &\left(\frac{c(1+\lambda)^2}{b}\left(\frac{v-2}{v-1}\right)\beta^{(v-1)/2}\right. \\ &\left. - \frac{(1+\lambda)a}{b}\left(1 - F_s\left(\frac{bq_\epsilon(\alpha)+a}{1+\lambda}\sqrt{\frac{v}{v-2}}, v\right)\right)\right) \end{aligned} \tag{2.5}$$

where $\beta = \left(cos\left(arctan\left(\dfrac{bq_\epsilon(\alpha)+a}{(1+\lambda)\sqrt{v-2}}\right)\right)\right)^2$, F_s is the CDF of a random variable with a symmetric Student-t distribution, v degrees of freedom, and α confidence level; and F is the CDF of a random variable with a skewed Student-t distribution and v degrees of freedom with skewness parameter λ.

The above DGP exhibits many of the stylized regularities observed in returns on financial assets, including asymmetric conditional variance with greater volatility for large negative returns and less volatility for positive returns (Hafner (1998)), long memory in volatility, significant collinearity with exogenous variables, conditional skewness (Patton (2004); Chen (2001); Ait-Sahalia and Brandt (2001)), leptokurdicity (Tauchen (2001); Andreou et al. (2001)), and nonlinear temporal dependence (Martins-Filho and Yao (2006)). The DGP is therefore able to adequately demonstrate most of the properties of financial returns and provides a useful approximation for our Monte Carlo.

2.2 Definitions of VaR and ES

Using the conventions of Martins-Filho and Yao (2006), VaR is formally defined as follows. Let $\{Y_t\}$ be a stochastic process representing a sequence of returns on a given financial asset, with discrete-time index t. Let the unknown conditional distribution of Y_t be denoted by F_t, which is absolutely continuous. F_t is conditioned on a sequence of lagged realizations, given as, $\{Y_{t-k}\}_{1\leq k\leq M}$, for some $M \geq 1$. For $0<\alpha<1$, the α-VaR of Y_t is the α-quantile of the conditional CDF, F_t. We denote it by $F_t^{-1}(\alpha|\{Y_{t-k}\}_{1\leq k\leq M})$ and assume that

$$F_t^{-1}(\alpha|\{Y_{t-k}\}_{1\leq k\leq M}) = \mu_t + \sigma_t q_\epsilon(\alpha) \tag{2.6}$$

Expressed informally, VaR gives the maximum financial loss on a portfolio over a given time horizon that will happen with probability not exceeding $1-\alpha$.

Expected shortfall is defined as $E_{F_t^y}(Y_t)$, which denotes the expected value taken with respect to F_t^y, the truncated distribution defined such that $Y_t>y$ where y is a specified loss threshold. Whenever the threshold y is taken to be α-VaR, we refer to α-ES. Expressed mathematically, expected shortfall is given as in Martins-Filho and Yao (2006) as,

$$E(Y_t|Y_t > F^{-1}(\alpha|\{Y_{t-k}\}_{1\leq k\leq M})) = \mu_t + \sigma_t E(\epsilon_t|\epsilon_t > q_\epsilon(\alpha)) \tag{2.7}$$

Informally, ES gives the expected loss on a financial asset or portfolio given that losses exceed a specified quantile.

Accurately estimating VaR and ES depends crucially on the ability to estimate the tails of the probability density function (PDF) f_t associated with the cumulative distribution function (CDF) F_t. Traditional methods of tail estimation are insufficient to accurately model tail events since the vast majority of realizations of the relevant random variable will take values near the center of the distribution (Diebold et al. (1993)). Extreme value theory (EVT) attempts to model the probability distributions of highly unlikely occurrences by approximating only the tails of f_t via an appropriately defined parametric density function. We discuss this further in section 2.5.

2.3 L-Moments and Maximum Likelihood Estimation

L-moments estimators are defined as summary statistics for probability distributions and data samples (IBM Corporation (2003)). L-moments are analogous to traditional moments in that they provide measures of location, dispersion, skewness, kurtosis, and higher-order moments for any probability distribution. L-moment estimators are computed using linear combinations of the ordered values of the data sample (Hosking (1990)).

Hosking (1990) outlined the following advantages of L-moments over conventional statistical moments :

- The probability distribution of the data sample must possess a finite mean, but need not possess any finite higher order moments. A distribution can be characterized uniquely by its L-moments as long as this is true (Martins-Filho and Yao (2006)).

- Sample L-moment ratios (analogous to standardized moments) can assume any value possible within the corresponding population.

- Asymptotic approximations of sampling distributions are better for L-moments than conventional moments (IBM Corporation (2003)).

- As a result of their definition as linear combinations of the data, L-moments are less susceptible to the effects of sampling variability and outliers in the data sample (Royston (1992)).

- L-moments allow for better inferences to be made from small samples about the probability distribution underlying the data sample.

- L-moments outperform ML estimators on an MSE basis in finite samples (Martins-Filho and Yao (2006); Hosking et al. (1985); Hosking and Wallis (1987)).

For a detailed treatment of the mathematical properties underlying the above claims, consult Hosking (1990); Hosking and Wallis (1997); Martins-Filho and Yao (2006).

We formally define L-moments both generally and for finite samples as they are presented in Martins-Filho and Yao (2006). Let ϵ be a random variable representing residuals and let F_ϵ be its CDF. Let $\alpha \in (0,1)$ and define $q_\epsilon(\alpha)$ as its quantile. For $r \in \mathbb{N}$, the r^{th} L-moment of ϵ is defined as,

$$\lambda_r = \int_0^1 q_\epsilon(\alpha) P_{r-1}(\alpha) d\alpha \tag{2.8}$$

where $P_r(\alpha) = \sum_{k=0}^{r} p_{r,k}\alpha^k$ and $p_{r,k} = \frac{(-1)^{r-k}(r+k)!}{(k!)^2(r-k)!}$. $P_r(\alpha)$ is the r^{th} shifted Legendre orthogonal polynomial. Conversely, conventional moments are defined by $\mu_r = \int_0^1 q_\epsilon(\alpha)^r d\alpha$, where μ_r is the general term for the r^{th} conventional moment.

L-moments can be used to estimate a finite number of parameters $\theta \in \Theta$, which characterize a member of a family of distributions. For $p \in \mathbb{N}$, let $\{F_\epsilon(\theta) : \theta \in \Theta \subset \mathbb{R}^p\}$ be a family of distributions known up to θ parameters. We denote our collection of residuals by $\{\epsilon_t\}_{t=1}^T$ where T is the size of the sample. As shown above, the L-moments, λ_r, uniquely characterize F_ϵ. This implies that θ may be expressed as a function of λ_r. If we are able to estimate $\hat{\lambda}_r$ from $\{\epsilon_t\}_{t=1}^T$, then we may also estimate $\hat{\theta}(\hat{\lambda}_1, \hat{\lambda}_2, ...)$. By equation (2.8), $\lambda_{r+1} = \sum_{k=0}^{r} p_{r,k}\beta_k$ for $r = 0, 1, ...$ where $\beta_k = \int_0^1 q_\epsilon(\alpha)\alpha^k d\alpha$ for $r = 0, 1, ...$ are the probability weighted moments. For $\{\epsilon_t\}_{t=1}^T$, we define $\epsilon_{(k)}$ as the k^{th} smallest element in the sample such that $\epsilon_{(1)} \leq \epsilon_{(2)} \leq ... \leq \epsilon_{(T)}$. As defined in Martins-Filho and Yao (2006), an unbiased estimator of β_k is

$$\hat{\beta}_k = \frac{1}{T}\sum_{j=k+1}^{T} \frac{(j-1)(j-2)...(j-k)}{(T-1)(T-2)...(T-k)}\epsilon_{(j)} \tag{2.9}$$

and we define $\hat{\lambda}_{r+1} = \sum_{k=0}^{r} p_{r,k}\hat{\beta}_k$ for $r = 0, 1, ..., T-1$.

One can also consider a different calculation methodology for L-moments in finite samples, as given by Wang (1997). Wang (1997) showed that the first four L-moments in a finite sample of data $x_{(t)}$ sorted into its order statistics, denoted $\lambda_1, \lambda_2, \lambda_3,$ and λ_4, can be expressed by,

$$\begin{aligned}
\lambda_1 &= \binom{T}{1}^{-1}\sum_{t=1}^{T} x_{(t)} \\
&= \frac{1}{T}\sum_{t=1}^{T} x_{(t)} && (2.10) \\
\lambda_2 &= \frac{1}{2}\binom{T}{2}^{-1}\sum_{t=1}^{T}\left\{\binom{t-1}{1} - \binom{T-t}{1}\right\} x_{(t)} \\
&= \frac{1}{T(T-1)}\sum_{t=1}^{T}(2t-T-1)x_{(t)} && (2.11) \\
\lambda_3 &= \frac{1}{3}\binom{T}{3}^{-1}\sum_{t=1}^{T}\left\{\binom{t-1}{2} - 2\binom{t-1}{1}\binom{T-t}{1} - \binom{T-t}{2}\right\} x_{(t)} \\
&= \frac{1}{T(T-1)(T-2)}\sum_{t=1}^{T}\left[(t-1)(t-2) - 4(t-1)(T-t) + (T-t)(T-t-1)\right] x_{(t)} && (2.12) \\
\lambda_4 &= \frac{1}{4}\binom{T}{4}^{-1}\sum_{t=1}^{T}\left\{\binom{t-1}{3} - 3\binom{t-1}{2}\binom{T-t}{1} - 3\binom{t-1}{1}\binom{T-t}{2} - \binom{T-t}{3}\right\} x_{(t)} \\
&= \frac{1}{T(T-1)(T-2)(T-3)}\sum_{t=1}^{T}\left[(t-1)(t-2)(t-3) - 9(t-1)(t-2)(T-t)\right. \\
&\quad \left. +9(t-1)(T-t)(T-t-1) - (T-t)(T-t-1)(T-t-2)\right] x_{(t)} && (2.13)
\end{aligned}$$

where $\binom{a}{b}$ is the binomial coefficient. We utilize the simplified versions of the first two L-moments in our procedure.

Assuming it exists, the first L-moment is a measure of the location of a distribution. The first L-moment is equivalent to the conventional first moment (i.e. $\lambda_1 = \mu_1$). λ_2 is a measure of the dispersion of the distribution and is a scalar multiple of the expectation of Gini's mean difference statistic.[1] λ_2 places smaller weights on the differences between estimates and realizations of the random variable and as such, it produces a measure of scale not equivalent to conventional variance (Hosking (1990)). Higher order moments are characterized as L-moment ratios, where for $r \in \mathbb{N}$, $\tau_r = \dfrac{\lambda_r}{\lambda_2}$. Therefore, L-skewness, the third moment, is denoted $\tau_3 \equiv \dfrac{\lambda_3}{\lambda_2}$. If μ_1 exists, $-1 < \tau_3 < 1$ with $\tau_3 = 0$ for symmetric distributions (Hosking (1989)). This means that L-skewness is bounded and therefore less sensitive to extreme values in the tails of the distribution than conventional, unbounded skewness. A similar result is observed by Oja (1981) for L-kurtosis, τ_4, where $-1 < \tau_4 < 1$. L-kurtosis is also bounded and less sensitive to outliers in the distribution. These characteristics of L-moments are desirable for modeling the statistical regularities present in financial time series, which are discussed in detail in section 2.6.

Our assumption that the tails of the distribution may be approximated by a generalized pareto distribution (discussed later) may be restrictive in MLE. If the tail is actually not prescribed by a GPD and the ML estimators are calculated under the assumption that it is, then the ML estimators may be biased, while the non-parametric L-moments may provide better estimates. Additionally, in "highly nonlinear dynamic models with fat tails and latent variables, asymptotic efficiency of the maximum likelihood (ML) estimator is not always warranted" (Andersen et al. (2009)). In fact,

[1] Gini's mean difference statistic is a measure of statistical dispersion that considers the average absolute difference between two realizations of a random variable drawn from a specified probability distribution. Provided n realizations of some random variable x, the mean difference is given by, $MD = \dfrac{\sum_{i=1}^{T}\sum_{j=1}^{T}|x_i - x_j|}{T(T-1)}$.

in nonstationary cases such as financial time series, the ML estimates are no longer asymptotically normal (Chan and Wei (1988); Phillips and Yu (2009)).

Maximum likelihood estimation (MLE) is the most popular method used by econometricians and statisticians to estimate the parameters of a model. Generally, for a set of data with an underlying probability distribution, MLE selects values for the parameters of the model that produce the distribution most likely to have generated the observed data. If the data are independent and identically distributed (IID), it is possible to express the joint density function $f(y_1, ..., y_T|\theta)$ where $y_1, ..., y_T$ are the observed values of the data and θ is a vector of parameters of the data. The joint density is therefore given by $f(y_1, ..., y_T|\theta) = f_1(y_1|\theta) \times ... \times f_T(y_T|\theta)$. If the data exhibit dependence, we may define the conditional joint density as $f(y_1, ..., y_T|\theta) = f(y_T|y_{T-1}, y_{T-2}, ..., y_1, \theta) \times ... \times f(y_1|\theta)$. Many practitioners use MLE to estimate model parameters because ML estimators exhibit several attractive asymptotic properties in stationary dynamic models (Wald (1949); Andersen et al. (2009); Hall and Heyde (1980); Billingsley (1961); Dacunha-Castelle and Florens-Zmirou (1986)), namely:

- Consistency - As the number of observations, T, grows, a sequence of ML estimators converges in probability to the true value ($\hat{\theta}_{mle} \xrightarrow{p} \theta_0$)

- Asymptotic normality - As T grows, the ML estimator assumes an asymptotically normal distribution when suitably standardized ($\sqrt{T}(\hat{\theta}_{mle} - \theta_0) \xrightarrow{d} N(0, I^{-1})$) where I is the Fisher Information matrix. For large T, ML estimators achieve the Cramer-Rao lower bound, meaning that there exists no asymptotically unbiased estimator with lower mean squared error (MSE).

Note that all of the above properties hold asymptotically. For the purposes of our paper, we seek to uncover the finite sample properties of our estimators for VaR and ES since all applied financial work is conducted using finite samples. As such, we employ L-moments estimation of the parameters of our DGP.

2.4 Local Polynomial Regression

Local polynomial regression is a nonparametric nonlinear estimation procedure that fits a regression function to a data series in a piecewise manner, considering only partial windows of the sample data at a time. We define local polynomial regression as in Fan (1992). Consider a sequence $(X_1, Y_1), ..., (X_T, Y_T)$ of random variables from a population with unknown density $f(x, y)$. The marginal density of X is therefore $f_X(x)$. The regression function $m(x)$ is a conditional expectation for Y, denoted $m(x) = E(Y_t|X_t = x)\ \forall t$. The conditional variance is given as $\sigma^2(x) = Var(Y_t|X_t = x)\ \forall t$.

Martins-Filho and Saraiva (2011) define local polynomial smoothers for univariate regressions as follows. Let a p^{th} order local polynomial regression estimator for conditional expectation of Y_t given regressor X_t denoted by $\hat{m}(x)$, be given by,

$$\hat{m}(x) \equiv (\hat{a}_{T0}(x;h), ..., \hat{a}_{Tp}(x;h)) = \underset{a_0,...,a_p}{\text{argmin}} \left(\sum_{t=1}^{T} \left(Y_t - \sum_{j=0}^{p} a_j (X_t - x)^j \right)^2 K\left(\frac{X_t - x}{h}\right) \right) \quad (2.14)$$

where K is a kernel estimator with optimally determined bandwidth h and vanishing higher-order moments. Higher order estimators for $m(x)$ are rarely used in practice because as the order p increases, the necessary assumption of p-times differentiability may become restrictive. For our first stage estimation procedure, we utilize local linear regression, which is the special case of equation (2.14) where $p = 1$.

2.5 Extreme Value Theory

Extreme value theory (EVT) is a field of statistics concerned with modeling maxima and extreme values of random variables. There are two traditional methods by which extreme values are modeled.

The cornerstone of EVT is the Fisher-Tippett-Gnedenko Theorem, which states that the maximum of a sample of IID random variables converges in distribution to one of only three possible families of distributions: the Gumbel distribution, the Frechet distribution, or the reverse Weibull distribution (Fisher and Tippett (1928)). Gnedenko (1943) later proved the necessary and sufficient conditions for which this result holds. These distributions are special cases of the generalized extreme value (GEV) distribution (McFadden (1978)).

An interesting result was obtained by Pickands (1975), which states that the distribution of the exceedances (residuals) of a random variable, ϵ, over a specified threshold, u, can be approximated by a generalized pareto distribution (GPD) with mean zero, provided F_ϵ belong to the domain of attraction of a Gumbel, Frechet, or reverse Weibull distribution. Let ϵ be a stochastic variable with shape parameter ψ and scale parameter β. The CDF of the GPD is given by,

$$F(\epsilon; \psi, \beta) = 1 - \left(1 + \psi \frac{\epsilon}{\beta}\right)^{-1/\psi}, \ \epsilon \in D$$

and the PDF is given by,

$$f(\epsilon; \psi, \beta) = \left(\frac{1}{\beta}\right)\left(1 + \psi \frac{\epsilon}{\beta}\right)^{-(1+\frac{1}{\psi})}$$

where $D = [0, \infty)$ if $\psi \geq 0$ and $D = [0, -\beta/\psi]$ if $\psi < 0$. Our estimation procedure makes use of this result to approximate only the tails, or extreme values, of the distribution underlying our data.

2.6 Properties of Financial Return Series and Modeling

2.6.1 Properties of Financial Time Series

Asymmetry of the conditional return distribution - Returns on financial assets exhibit leptokurtosis, meaning that their probability distributions possess 'fat tails.' From a modeling perspective, fat tails imply that there is a greater probability of experiencing large gains or losses than under the assumption of normality. As such, modeling procedures employing an assumption of normality ignore the significant impact of higher-order moments on their estimation, particularly in tail estimation.

Financial returns also exhibit negative conditional skewness, meaning a larger portion of the probability density function takes values below the median than a symmetric distribution would predict. With respect to returns modeling, ignoring skewness will overestimate the likelihood of large positive gains while also underestimating the likelihood of large losses. For more thorough characterizations of these results, see Tauchen (2001); Andreou et al. (2001); Hafner (1998); Ait-Sahalia and Brandt (2001); Chen (2001); Patton (2004); Gallant and Tauchen (1989), and Bodie et al. (2009).

Asymmetric conditional volatility - Literature suggests that volatility in returns of a financial asset tends to be greater in a downward trend than in an upward trend. Essentially, when losing value, we tend to see more volatility than when gaining value. This property has particularly salient applications to event studies, such as those dealing with the impact of news on returns, in which negative news releases typically have a greater impact than positive releases (Engle and Ng (1993)). One need only examine the historical record to see far greater volatility in periods of recession and economic contraction than in periods of expansion. For more thorough characterizations of this result, see Kroner and Ng (1998); Black (1976); Pagan and Schwert (1990); Engle and Ng (1993), and Hafner (1998).

Long memory in returns - The literature suggests that financial time series exhibit long memory in returns, meaning returns exhibit high autocorrelations with prior returns. This result essentially negates the assumption that financial return series are independent. Characteristics of the markets or assets being analyzed do have a significant impact on long memory properties. Limam (2003) found, for example, that long memory tends to exist more in thin markets. Long memory in returns is observed much less often in very liquid markets.

Specifically, financial time series are characterized by nonlinear temporal dependence (Martins-Filho and Yao (2006)). Correlogram plots for various financial time series show a distinct hyperbolic decay in correlation that is well-described by a fractionally-integrated process (Andersen et al. (2009)). Similar results are found for currencies (Andersen and Bollerslev (1997, 1998); Andersen et al. (2001); Cheung (1993); Zumbach (2004)), equities (Andersen et al. (2001); Areal and Taylor (2002); Deo et al. (2006); Martens (2002)), and bond yields (Andersen and Benzoni (2010)). For more results concerning long memory in returns, see Breidt et al. (1998); Engle and Lee (1999); Goetzmann (1993); Nawrocki (1993), and Huang and Yang (1999).[2]

Volatility clustering - Volatility clustering means that periods of high volatility tend to follow periods of high volatility, while periods of low volatility tend to follow periods of low volatility. This stylized fact is one of the most persistent and is most consistently supported in the literature. Volatility persistence is the econometric analog to Newton's First Law of Motion: an object at rest tends to remain at rest and an object in motion tends to remain in motion. This property is also known as volatility persistence, long memory in volatility, or serial correlation in volatility. For more thorough characterizations of these results, see Fan and Yao (1998); Bollerslev et al. (1992); Bollerslev (1986); Mandelbrot (1963); Fama (1965); Poterba and Summers (1986); Engle and Mustafa (1992), and Milhoj (1985).

Leverage effect - A reduction in the equity value of a financial asset would raise its debt-to-equity ratio, implying greater riskiness of the asset in the form of an increase in future volatility. As a result, future volatility is negatively related to the current return on a financial asset. This property is known as the leverage effect. For more information on the leverage effect, consult Black

[2]For papers which find do not find support for the property of long memory in returns, see Lo (1991); Lobato and Savin (1998); Oh et al. (2006); Chow et al. (1996); Grau-Carles (2005)

(1976); Christie (1982); Kupiec (1989); Chen (2001); Ding et al. (1993); Hafner (1998); Engle and Patton (2001); Patton (2004), and Gallant et al. (1992).

2.6.2 Modeling Financial Returns

This subsection provides a brief discussion of the statistical approaches used to model returns on financial assets. It serves as a historical account of the developments leading to the modeling techniques employed both in recent literature and this paper. Note that the following discussion is by no means exhaustive or comprehensive. For more thorough treatments of time series modeling, see Andersen et al. (2009); Terasvirta (2008).

ARMA Models - Autoregressive Moving Average (ARMA) models are a class of time series models used to analyze and forecast stationary stochastic time indexed variables. We let $\{Y_t\}$ be a univariate covariance stationary time series. The property of stationarity requires that $E[Y_t]$ is independent of the time index t and that $Cov(Y_t, Y_{t+h})$ is finite and depends only on the lag, h, not the position in the series. We define ARMA models as in Holan et al. (2010). The series $\{Y_t\}$ is an ARMA series with autoregressive order $p \geq 0$ and moving average order $q \geq 0$ if it is stationary and a solution to the equation given by,

$$Y_t = \alpha + \sum_{i=1}^{p} \phi_i Y_{t-i} + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} + \epsilon_t \tag{2.15}$$

where α is some intercept constant and $\{\epsilon_t\}$ is a mean zero white noise (IID) process of residuals with $Var(\epsilon_t) \equiv \sigma_t^2$. The parameters of the ARMA model are estimated by MLE, Method of Moments, or OLS regression. For clarity, autoregressive terms are previous realizations of the regressand, while the moving average terms are past realizations of error terms, which are typically assumed to follow a prescribed distribution, Gaussian or otherwise.

ARMA models may be generalized to include the effects of other exogenous variables, as is the case in the autoregressive moving average with exogenous inputs (ARMAX) model (Peng et al.

(2001)). The ARMAX model is defined as,

$$Y_t = \alpha + \sum_{i=1}^{p} \phi_i Y_{t-i} + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} + \sum_{i=1}^{b} \eta_i d_{t-i} + \epsilon_t \tag{2.16}$$

where $\{d_t\}$ is a exogenous time series and $\eta_1, ..., \eta_b$ are the coefficients of $\{d_t\}$. ϵ_t and α are defined as above. This is the alternative first stage estimation approach used in our Monte Carlo simulation.

Even more general is the nonlinear autoregressive exogenous (NARX) model presented by Leontaritis and Billings (1985a,b), which is given by,

$$Y_t = m(Y_{t-1}, Y_{t-2}, ..., d_t, d_{t-1}, d_{t-2}, ...) + \epsilon_t \tag{2.17}$$

where ϵ_t remains the white noise error term and m is some nonlinear function estimated through nonlinear regression techniques or machine learning algorithms.

ARCH/GARCH Models[3] **-** The class of autoregressive conditionally heteroskedastic (ARCH) models arose to address the property of serial correlation and non-stationarity in asset returns. Engle (1982) introduced the ARCH(q) model where conditional variance is written as a distributed lag of q past squared innovations,

$$\sigma_t^2 = \alpha + \sum_{i=1}^{q} \beta_i \epsilon_{t-i}^2 \tag{2.18}$$

where α is the intercept parameter and the β_i are the coefficients on the lagged residuals, ϵ_{t-i}, which are assumed to be distributed $N(0, \sigma^2)$. For the conditional volatility to be positive, note that the α and β_i coefficients must also be positive.

Bollerslev (1986) later proposed the generalized autoregressive conditionally heteroskedastic (GARCH) model to reduce the number of β_i coefficients while still capturing persistence in volatility. The GARCH(p, q) model, where p is the order of the autoregressive lags on the σ_{t-i}^2 terms and q is

[3] For all the GARCH definitions, we define $\epsilon_t = \sigma_t z_t$ where σ_t is the standard deviation of the data and z_t is an error process following a prescribed distribution, typically IID standard normal.

the order of the moving average lags on the ϵ_{t-i}^2 terms, is given as,

$$\sigma_t^2 = \alpha + \sum_{i=1}^{p} \omega_i \sigma_{t-i}^2 + \sum_{i=1}^{q} \beta_i \epsilon_{t-i}^2 \tag{2.19}$$

where α is the intercept parameter, the ω_is are the coefficients on the autoregressive lags, and the β_is are the coefficients on the moving average lags. Though the GARCH model is also a weighted average of past squared residuals, it is different because it contains declining weights that never equal zero. This characteristic captures the persistent memory documented in asset returns. Note that "GARCH models are mean reverting and conditionally heteroskedastic, but have a common unconditional variance" (Sheth and Kim (2003)).

In 1993, Engle and Ng introduced the nonlinear GARCH (NGARCH) model to capture asymmetry. NGARCH is a special case of the GARCH(1,1) model, given by,

$$\sigma_t^2 = \alpha + \beta(\epsilon_{t-1} - \theta\sigma_{t-1})^2 + \omega\sigma_{t-1}^2 \tag{2.20}$$

where $\beta, \omega \geq 0$ and $\alpha > 0$. NGARCH models demonstrate the leverage effect when θ is estimated to be positive (Posedel (2006)).

The exponential GARCH (EGARCH) model proposed by Nelson (1990) is a nonlinear expansion of the traditional GARCH model. Nelson defined σ_t^2 as an asymmetric function of past residuals, ϵ_t, given by,

$$ln(\sigma_t^2) = \alpha_0 + \sum_{i=1}^{p} \alpha_i \left(\phi z_{t-i} + \gamma(|z_{t-i}| - E|z_{t-i}|)\right) + \sum_{i=1}^{q} \beta_i ln(\sigma_{t-i}^2) \tag{2.21}$$

where the error process z_t is assumed IID with mean zero and unit variance. The coefficients of this model are estimated by maximum likelihood. Where the EGARCH(p, q) model departs from the classical GARCH model is its lack of restrictions on α_i and β_i. Such restrictions serve to ensure non-negativity of the conditional variances in the GARCH model but are unnecessary in the EGARCH model.

The EGARCH model captures leverage effects and the asymmetric conditional volatility noted by Black (1976) and others. In the EGARCH model, if $\alpha_i \phi < 0$, the variance will increase when $\epsilon_{t-i} < 0$ and vice versa. The EGARCH model also allows for "random oscillatory behavior in the σ_t^2 process"(Sheth and Kim (2003)). The absence of restrictions on the β_i terms allows oscillations since the coefficients can be positive or negative. A benefit of this approach, noted by Campbell et al. (1997), is that it does not require parametric restrictions for the conditional variance to be positive. Moreover, in the special case that $\alpha + \beta = 1$, the EGARCH model is both strictly nonstationary and covariance stationary (Sheth and Kim (2003)). Finally, the EGARCH model is more robust to extreme shocks than traditional GARCH models.

The final GARCH procedure we outline is the threshold GARCH (TGARCH) model of Zakoian (1994). Rather than using squared residuals like most other GARCH variants, the TGARCH method uses absolute residuals. This is done because Davidian and Carroll (1987) found that absolute residuals yield more efficient variance estimates under non-normal distributions than squared residuals. Therefore, the TGARCH model specifies a conditional standard deviation rather than a conditional variance. What distinguishes the TGARCH model is that the current volatility responds differently based on the sign of past innovations. Given that the residuals, denoted ϵ_t, we let $\epsilon_t^+ = max(\epsilon_t, 0)$ and $\epsilon_t^- = min(\epsilon_t, 0)$. The TGARCH$(p, q)$ process is given as,

$$\sigma_t = \alpha_0 + \sum_{i=1}^{p} (\alpha_i^+ \epsilon_{t-i}^+ - \alpha_i^- \epsilon_{t-i}^-) + \sum_{j=1}^{q} \beta_j \sigma_{t-j} \tag{2.22}$$

where ϵ_t is independent of Y_t and $\{\alpha_i^+\}_{i=1,\ldots,p}, \{\alpha_i^-\}_{i=1,\ldots,p}$, and $\{\beta_j\}_{j=1,\ldots,q}$ are real scalar sequences. If we do not assume σ_t is positive, then we must impose positivity constraints where $\alpha_0 > 0, \alpha_i^+ \geq 0, \alpha_i^- \geq 0$, and $\beta_i \geq 0 \ \forall i$. This model also captures asymmetrical conditional volatility and the leverage effect. As is summarized in Sheth and Kim (2003), the TGARCH model differs from the EGARCH model in several important aspects. For one, TGARCH is an additive model which makes volatility a function of non-normalized residuals. Furthermore, TGARCH allows for different lags to have opposite signs, while EGARCH imposes the same structure for all lags.

There exist a multitude of other ARCH/GARCH variants, each defining variance differently. For fairly comprehensive discussions of the assorted ARCH/GARCH models, please consult Andersen et al. (2009); Terasvirta (2008); Sheth and Kim (2003), and Bollerslev (2007).

CHARN Models - The conditional heteroskedastic autoregressive nonlinear (CHARN) model is a special case of the nonlinear-ARCH model considered by Masry (1995). The CHARN model is a nonlinear generalization of the GARCH(p,q) model expressed as a Markov chain, where m_t is a nonparametric function of $Y_{t-1}, ..., Y_{t-p}$ and σ_t is a nonlinear function of $Y_{t-1}, ..., Y_{t-q}$. This general CHARN process is given by

$$Y_t = m(Y_{t-1}, ..., Y_{t-p}) + \sigma(Y_{t-1}, ..., Y_{t-q})\epsilon_t \tag{2.23}$$

The CHARN process considered by Martins-Filho and Yao (2006); Diebolt and Guégan (1993); Hardle and Tsybakov (1997), and Hafner (1998) is expressed as a Markov chain of order 1 and is given by

$$Y_t = m(Y_{t-1}) + \sigma(Y_{t-1})\epsilon_t \text{ for } t = 1, 2, ... \tag{2.24}$$

where ϵ_t is an independent strictly stationary process with an unknown continuous marginal distribution F_ϵ with mean zero and unit variance. Assume ϵ_t is independent of all regressors. We assume skewness and kurtosis of F_ϵ exist, are continuous, and that m_t and σ_t^2 are twice differentiable. CHARN models capture the asymmetry in lagged values of Y_t that arises due to the leverage effect, which GARCH models fail to do. Martins-Filho and Yao (2006) do concede, however, that the CHARN model is more restrictive than GARCH models in that its Markov nature makes it less able to model the long memory property of asset return processes.

3 VaR and ES Estimation Method

The model we propose combines the flexibility of the MFY model with an exogenous variable as is done in the NARX model. We consider the following nonparametric definitions of μ_t, σ_t, and Y_t.

Assume $\{(Y_t, Y_{t-1}, D_{t-1})'\}$ is a 3-dimensional strictly stationary process with conditional mean function $E(Y_t|Y_{t-1} = x_1, D_{t-1} = d_1) = m(x_1, d_1)$ and conditional variance $E((Y_t - m(x_1, d_1))^2|Y_{t-1} = x_1, D_{t-1} = d_1) = \sigma^2(x_1, d_1) > 0$ where D_{t-1} represents a one-period lagged exogenous variable. For $t \in \mathbb{N}$, the process is described by,

$$Y_t = m(Y_{t-1}, D_{t-1}) + \sigma(Y_{t-1}, D_{t-1})\epsilon_t \tag{3.1}$$

where ϵ_t are independent, strictly stationary residuals with an unknown absolutely continuous marginal distribution function F_ϵ with mean zero and unit variance. Assume ϵ_t is independent of both Y_{t-1} and D_{t-1}. Assume conditional skewness, $E(\epsilon_t^3)$, and kurtosis, $E(\epsilon_t^4)$, exist and are continuous. Further assume that $m(x_1, d_1)$ and $\sigma^2(x_1, d_1)$ are twice differentiable on the open set containing x_1 and d_1. Unfortunately, the estimators for m and σ^2 in the nonparametric generalization of ARCH and GARCH (1,1) models proposed by Carroll et al. (2002) converge exponentially more slowly as the number of lags in the conditioning set increases, which is the curse of dimensionality. Since our model incorporates an exogenous variable, it may be more susceptible to the curse of dimensionality than the model proposed by Martins-Filho and Yao (2006). This indicates that a high number of observations are necessary to obtain adequate asymptotic approximations.

3.1 Estimation of $\hat{m}$ and $\hat{\sigma}^2$

In our estimation of m and σ^2, we consider the estimation procedure first proposed by Fan and Yao (1998) and later used by Martins-Filho and Yao (2006), but generalize it to the multivariate case. Let $\mathbf{X}$ be a matrix of the regressors considered, namely $[Y_{t-1}D_{t-1}]_{t=1,\ldots,T}$, where D_{t-1} is an one-period lagged exogenous variable behaving like the S&P 500. We estimate $m(\mathbf{X})$ and $\sigma^2(\mathbf{X})$ with a generalized version of the procedure given by Fan (1992), which is detailed in the literature review. Since our process is defined in equation (3.1) as a function of Y_{t-1} and D_{t-1}, we cannot use the univariate local linear regression model shown in the literature review.

In the univariate case, suppose we have the local linear regression given by,

$$\hat{m}(x) = \underset{a_0,a_1}{\operatorname{argmin}} \sum_{t=1}^{T} \left(Y_t - a_0 - a_1\left(X_t - x\right)\right)^2 K\left(\frac{X_t - x}{h}\right)$$

We can then express the same function in terms of vectors instead of in summation notation. Rewriting in this manner yields,

$$\hat{m}(x) = \underset{a_0,a_1}{\operatorname{argmin}}[\mathbf{Y} - \mathbf{1_T}\mathrm{a}_0 - \mathrm{a}_1(\mathbf{X} - \mathbf{1_T}\mathrm{x})]'\mathbf{K}[\mathbf{Y} - \mathbf{1_T}\mathrm{a}_0 - \mathrm{a}_1(\mathbf{X} - \mathbf{1_T}\mathrm{x})] \tag{3.2}$$

where $\mathbf{X}$ and $\mathbf{Y}$ are $T \times 1$ vectors of dependent and independent variables, respectively, $\mathbf{K}$ is a diagonal matrix with dimension $T \times T$ whose diagonal elements are the kernel evaluated at $\dfrac{X_t - x}{h}$ and $\mathbf{1_T}$ is a column vector of ones with dimension $T \times 1$. It is easily verified that

$$\hat{m}(x) = \underset{a_0,a_1}{\operatorname{argmin}} \left(\mathbf{Y} - \left[\begin{array}{cc} \mathbf{1_T} & (\mathbf{X} - \mathbf{1_T}\mathrm{x}) \end{array}\right] \left[\begin{array}{c} a_0 \\ a_1 \end{array}\right]\right)' \mathbf{K} \left(\mathbf{Y} - \left[\begin{array}{cc} \mathbf{1_T} & (\mathbf{X} - \mathbf{1_T}\mathrm{x}) \end{array}\right] \left[\begin{array}{c} a_0 \\ a_1 \end{array}\right]\right) \tag{3.3}$$

If we define $R \equiv \left[\begin{array}{cc} \mathbf{1_T} & (\mathbf{X} - \mathbf{1_T}\mathrm{x}) \end{array}\right]$ and $\gamma \equiv \left[\begin{array}{c} a_0 \\ a_1 \end{array}\right]$, then

$$\hat{m}(x) = \underset{\gamma}{\operatorname{argmin}}(\mathbf{Y} - \mathbf{R}\gamma)'\mathbf{K}(\mathbf{Y} - \mathbf{R}\gamma)$$

with solution

$$\hat{\gamma} = (\mathbf{R}'\mathbf{K}\mathbf{R})^{-1}\mathbf{R}'\mathbf{K}\mathbf{Y} \tag{3.4}$$

Now we generalize to L regressors. In our case, where there is one exogenous variable, our local linear regression will have $L = 2$ regressors since m and σ^2 are functions of Y_{t-1}, D_{t-1} for $t = 1, ..., T$. We will therefore express the estimator for $m(\mathbf{x})$ as

$$\hat{m}(\mathbf{x}) = \underset{\gamma_\mathrm{a}}{\operatorname{argmin}}(\mathbf{Y} - \mathbf{R}\gamma_\mathbf{a})'\mathbf{K}(\mathbf{Y} - \mathbf{R}\gamma_\mathbf{a}) \tag{3.5}$$

where

$$\mathbf{R} = \begin{bmatrix} \mathbf{1_T} & (\mathbf{X_1} - \mathbf{1_T}\mathrm{x}_1) & \ldots & (\mathbf{X_L} - \mathbf{1_T}\mathrm{x}_\mathrm{L}) \end{bmatrix} \tag{3.6}$$

and

$$\hat{\gamma}_\mathbf{a} = \begin{bmatrix} \hat{a}_0 \\ \vdots \\ \hat{a}_L \end{bmatrix} = (\mathbf{R}'\mathbf{K}\mathbf{R})^{-1}\mathbf{R}'\mathbf{K}\mathbf{Y} \tag{3.7}$$

We are only concerned with a_0, so we multiply $\hat{\gamma}_\mathbf{a}$ by $\mathbf{e}$, a $1 \times L+1$ row vector with first element one and all other elements zero. Hence,

$$\hat{m}(\mathbf{x}) = \mathbf{e}\hat{\gamma}_\mathbf{a} \tag{3.8}$$

We should also note that the definition of our kernel function, K, is a multiplicative kernel. Since we utilize an IID standard normal kernel function, our kernel becomes a multivariate standard normal density. Our new multiplicative kernel, $\mathbf{K}(\mathbf{l})$ for $l = 1, ..., L$, is then a diagonal matrix given by,

$$\mathbf{K} = \mathrm{diag}\left\{ \prod_{\mathrm{l}=1}^{\mathrm{L}} \mathrm{K}\left(\frac{\mathrm{X}_{\mathrm{t,l}} - \mathrm{x}_\mathrm{l}}{\mathrm{h}_{0\mathrm{l}}} \right) \right\}_{\mathrm{t}=1,\ldots,\mathrm{T}} \tag{3.9}$$

where each regressor has its own bandwidth, h_{0l}, for $l = 1, ..., L$ and $\mathbf{K}(\mathrm{l}) : \mathbb{R} \to \mathbb{R}$. We assume the bandwidths h_{0l} are sequences of positive real numbers such that $h_{0l} \to 0$ as $T \to \infty$.

Similarly, we define the local linear estimator of $\sigma^2(\mathbf{x})$ as,

$$\hat{\sigma}^2(\mathbf{x}) = \underset{\gamma_\mathrm{b}}{\mathrm{argmin}} (\hat{\mathbf{r}} - \mathbf{R}\gamma_\mathbf{b})'\mathbf{W}(\hat{\mathbf{r}} - \mathbf{R}\gamma_\mathbf{b}) \tag{3.10}$$

where the matrix of squared residuals, $\hat{\mathbf{r}}$, is defined as $\hat{\mathbf{r}} = (\mathbf{Y} - \hat{\mathrm{m}}(\mathbf{x}))^2$ for, $\mathbf{R}$ is defined as in equation (3.6), $\mathbf{W}$ is a multiplicative Gaussian kernel function characterized by,

$$\mathbf{W} = \mathrm{diag}\left\{ \prod_{\mathrm{l}=1}^{\mathrm{L}} \mathrm{W}\left(\frac{\mathrm{X}_{\mathrm{t,l}} - \mathrm{x}_\mathrm{l}}{\mathrm{h}_{1\mathrm{l}}} \right) \right\}_{\mathrm{t}=1,\ldots,\mathrm{T}} \tag{3.11}$$

and $\gamma_\mathbf{b}$ is given by,

$$\gamma_\mathbf{b} = \begin{bmatrix} b_0 \\ \vdots \\ b_L \end{bmatrix} \tag{3.12}$$

Similar to $\hat{m}(\mathbf{x})$, for $\hat{\sigma}(\mathbf{x})^2$ we are only concerned with b_0, so we multiply $\hat{\gamma_{\mathbf{b}}}$ by $\mathbf{e}$. That is, $\hat{\sigma}^2(x)$ is given by

$$\hat{\sigma}(\mathbf{x})^2 = \mathbf{e}\hat{\gamma}_{\mathbf{b}} \tag{3.13}$$

We estimate the sequences of bandwidths h_v using the empirical plug-in method proposed by Ruppert et al. (1995). The Ruppert bandwidth selection method is computationally superior to the cross-validation method and is a consistent estimator of the optimal bandwidth sequence that minimizes the asymptotic mean integrated squared error (MISE) of $\hat{m}$ and $\hat{\sigma}^2$ (Martins-Filho and Yao (2006)). The kernel function we use in our estimation is standard Gaussian, though many variants are available. See Li and Racine (2007) for a more thorough discussion of kernel estimators. $\hat{m}(\mathbf{X})$ and $\hat{\sigma}^2(\mathbf{X})$ are the first stage estimators for μ_t and σ_t^2, respectively, as seen in equations (2.6) and (2.7).

3.2 Estimation of β and ψ Using L-moments

To estimate β and ψ, we use Hosking's L-moments estimation procedure described in section 2.3. Martins-Filho and Yao (2006) showed that when the CDF F_ϵ is a GPD with the set of parameters $\theta = (\mu, \beta, \psi)$, then the location parameter $\mu = \lambda_1 - (2-\psi)\lambda_2$, the scale parameter $\beta = (1-\psi)(2-\psi)\lambda_2$, and the shape parameter $\psi = -\dfrac{1-3(\lambda_3/\lambda_2)}{1+(\lambda_3/\lambda_2)}$. For our purposes, $\mu = 0, \beta = (1-\psi)\lambda_1, \psi = 2 - \lambda_1/\lambda_2$. Therefore, the L-moment estimators for ψ and β are given by,

$$\hat{\psi} = 2 - \frac{\hat{\lambda}_1}{\hat{\lambda}_2} \tag{3.14}$$

$$\hat{\beta} = (1-\hat{\psi})\hat{\lambda}_1 \tag{3.15}$$

As Martins-Filho and Yao (2006) proved, our L-moment estimators are $\sqrt{T}$-asymptotically normal if $\psi < 0.5$. Our motivation for using L-moments despite the asymptotic efficiency of ML estimators is that they are much easier to compute than ML estimators because no iteration or optimization is necessary and because they may actually outperform ML estimators in finite samples (Hosking (1990)). Since our study is concerned only with the properties of the estimators in finite samples that may be too small for ML estimators to be used as proxies for the asymptotic distribution, we instead utilize L-moments.

3.3 Estimation of VaR and ES

The second stage of our estimation procedure provides estimators for $q_\epsilon(\alpha)$ and $E(\epsilon_t|\epsilon_t > q_\epsilon(\alpha))$ and subsequently VaR and ES. To conduct this estimation, we approximate the distribution of the exceedances, Z, where $Z = \epsilon - u$. The random variable ϵ represents a residual and u represents a specified threshold, as explained in section 2.5. This CDF and PDF of the GPD are included again here for reference.

$$F(\epsilon;\psi,\beta) = 1 - \left(1 + \psi\frac{\epsilon}{\beta}\right)^{-1/\psi}, \ \epsilon \in D$$
$$f(\epsilon;\psi,\beta) = \left(\frac{1}{\beta}\right)\left(1 + \psi\frac{\epsilon}{\beta}\right)^{-1-\frac{1}{\psi}}$$

where $D = [0,\infty)$ if if $\psi \geq 0$ and $D = [0, -\beta/\psi]$ if $\psi < 0$. Recall that ψ is the shape parameter and β is the scale parameter. We may then use the estimates of $\hat{\mu}_t$ and $\hat{\sigma}_t^2$ to generate a sequence of standardized residuals of the form $\left\{e_t = \dfrac{y_t - \hat{\mu}_t}{\hat{\sigma}_t}\right\}_{t=1}^{T}$. These residuals may then be used to estimate the tails of f_ϵ using the GPD. We first order the residuals from largest to smallest, where $e_{(j)}$ is the j^{th} largest residual. We fix a number, k, to be the number of residuals used in the estimation, which also implies a threshold, u. This threshold is defined as the $(k+1)^{th}$ largest residual such that $u = e_{(k+1)}$. We may then find $k < n$ exceedances over $e_{(k+1)}$ given by $\{e_{(j)} - e_{(k+1)}\}_{j=1}^{k}$. These excesses will then be used to estimate a GPD. Martins-Filho and Yao (2006) showed that for $\alpha > (1 - k/T)$, given estimates $\hat{\beta}$ and $\hat{\psi}$, we can estimate $q_\epsilon(\alpha)$ and $E(\epsilon_t|\epsilon_t > q_\epsilon(\alpha))$ by,

$$\hat{q}_\epsilon(\alpha) = e_{(k+1)} + \frac{\hat{\beta}}{\hat{\psi}}\left(\left(\frac{1-\alpha}{k/T}\right)^{-\hat{\psi}} - 1\right) \tag{3.16}$$

and for $\psi < 1$

$$\hat{E}(\epsilon_t|\epsilon_t > \hat{q}_\epsilon(\alpha)) = \hat{q}_\epsilon(\alpha)\left(\frac{1}{1-\hat{\psi}} + \frac{\hat{\beta} - \hat{\psi}e_{(k+1)}}{(1-\hat{\psi})\hat{q}_\epsilon(\alpha)}\right) \tag{3.17}$$

The specification of k is addressed in the Monte Carlo study in section 4 of this paper. Once we ascertain the estimators in equations (3.8) and (3.13), we may then use equations (2.6) and (2.7) to estimate $\alpha - VaR$ and $\alpha - ES$. These estimates are given by,

$$V\hat{a}R = \hat{F}^{-1}(\alpha|\mathbf{X}) = \hat{\mu}_{\mathrm{t}} + \hat{\sigma}_{\mathrm{t}}\hat{\mathrm{q}}_\epsilon(\alpha) \tag{3.18}$$

and

$$\hat{ES} = \hat{E}(Y_t|Y_t > \hat{F}^{-1}(\alpha|\mathbf{X}), \mathbf{X}) = \hat{\mu}_{\mathrm{t}} + \hat{\sigma}_{\mathrm{t}}\hat{\mathrm{E}}(\epsilon_{\mathrm{t}}|\epsilon_{\mathrm{t}} > \hat{\mathrm{q}}_\epsilon(\alpha)) \tag{3.19}$$

where Y_t is our time-indexed dependent variable and $\mathbf{X}$ is our matrix of regressors. Once we obtain these estimates, we have completed the one-period forecast for VaR and ES under α confidence for the regressand Y using the explanatory variables contained in $\mathbf{X}$.

3.4 Alternative First Stage Estimation Procedures

To compare the proposed first stage estimator given in section 3.1, we also consider the ARMAX OLS linear regression and GARCH method as an alternative first stage estimation procedure for m and σ^2. We regress the Y_t series on Y_{t-1} and D_{t-1} to obtain OLS estimates for the β coefficients in the following regression:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 D_{t-1} + \delta_t \tag{3.20}$$

where the δ_t are errors distributed with zero mean and variance σ^2. Using the estimated $\hat{\beta}_0, \hat{\beta}_1$, and $\hat{\beta}_2$, we can construct a series of squared residuals for the GARCH estimator, given by

$$\hat{\epsilon}_t^2 = (Y_t - (\hat{\beta}_0 + \hat{\beta}_1 Y_{t-1} + \hat{\beta}_2 D_{t-1}))^2 \tag{3.21}$$

Using ϵ_t, we perform the GARCH(1,1) procedure. Our GARCH method uses the series of squared residuals, $\hat{\epsilon}^2$, to yield the γ coefficients in the following model:

$$\sigma_t^2 = \gamma_0 + \gamma_1(\hat{\epsilon}_t^2) + \gamma_2\sigma_{t-1}^2 \tag{3.22}$$

Once estimated via maximum likelihood, we have the first stage GARCH estimators for m and σ^2. These estimates are then used in the second stage L-moments estimation and to produce estimates for VaR and ES. It is these GARCH estimates that form our benchmark in the Monte Carlo study.

We could also consider a much wider array of first stage estimation procedures, including several ARCH/GARCH variants, ARMA/ARIMA models, and more advanced modeling techniques. We limit ourselves to the ARMAX/GARCH estimator above because of its frequent use in empirical finance and ability to model some of the stylized facts about returns. For a more comprehensive treatment of the alternative procedures, see Andersen et al. (2009); Sheth and Kim (2003); Gouriéroux (1997); Teräsvirta and Zhao (2006). We leave a more thorough comparative study for future investigations.

4 Monte Carlo Simulation

In order to gain substantial insight into the properties of our proposed estimator, we designed a fairly comprehensive Monte Carlo simulation. The primary purpose of this simulation is to evaluate the relative performance of our estimators versus the frequently utilized GARCH(1,1) modeling technique. Secondarily, our Monte Carlo study provides researchers and practitioners with some guidance into selecting estimators for VaR and ES.

Similar to the approach utilized by Martins-Filho and Yao (2006), our data generating process (DGP), described in detail in section 2.1, is designed to capture the stylized facts about returns and volatility of financial assets. What differentiates this DGP from Martins-Filho and Yao (2006) and from other previous approaches, to our knowledge, is the consideration of an additional explanatory variable, D. To generalize the performance of our estimator in various conditions, our Monte Carlo simulation method afforded us the flexibility to examine results in a large number of varied conditions while also generating fairly large samples and producing numerous iterations of each scenario. To accomplish this, we varied the individual parameters of the model across 64 different scenarios, which are enumerated in Appendix A, Table 1.

The design of the Monte Carlo aims to yield relative performance metrics for our estimator and a GARCH(1,1) model in a variety of parameter configurations. We designed 64 experiments for the DGP. Table 1 in Appendix A provides the key for how the experiments are numbered in the results section. We consider the following parameter values:

- Two values for sample size: $n_S = \{500, 1000\}$
- Two values for λ: $n_\lambda = \{0, -0.5\}$
- Two values for γ: $n_\gamma = \{0.3, 0.9\}$
- Two values for the confidence level, α: $n_\alpha = \{0.95, 0.99\}$
- Two values for the number of exceedances, k: $n_k = \{60, 100\}$
- Two functional forms of $g(x_t)$, where x_t is a linear combination of the regressors, defined by $x_t = w_1 y_{t-1} + w_2 d_{t-1}$, are given by

  -From Hafner (1998), $g_1(x_t)$ is given by $g_1(x_t) = 0.5 + \dfrac{exp(-4x_t)}{1 + exp(-4x_t)}$

  -From Carroll et al. (2002), $g_2(x_t)$ is given by $g_2(x_t) = 1 - 0.9exp(-2x_t^2)$

The weights w_1 and w_2 are fixed at 0.4 and 0.3, respectively, throughout the Monte Carlo. This weighting system, while somewhat arbitrary, gives the most weight to the most recent observation and less weight to the exogenous variable. For our main returns series, the degrees of freedom parameter, v, is held constant at 8. For the exogenous variable, data were generated from a skewed Student-t distribution with $v = 3, \lambda = -0.1$, and $\gamma = 0.6$, which produces realizations that appear similar to the returns on the S&P 500 from January 3, 1950 to September 30, 2011 by inspection. For each of the 64 experiments, total trials were fixed at 500. Figure 1A in Appendix A illustrates the shapes of the $g_1(x_t)$ and $g_2(x_t)$ volatility functions.

In addition to the estimator considered above, we also performed the same Monte Carlo simulation routine for the MFY estimator. We did this to both attempt to recreate the results of Martins-Filho and Yao (2006) and to provide a basis of comparison for the multivariate estimator presented in this paper. Primarily, we are interested in the impact of adding a second regressor to the estimation procedure as it pertains to the performance metrics of root mean squared error (RMSE) and bias relative to the Gaussian GARCH estimator. We hope to gauge the impact of the curse of dimensionality, which can cripple the local linear estimator in small samples. The optimal rate of convergence of the local regression procedure decreases exponentially with the addition of each dimension (regressor) (Stone (1980)). Therefore, since the sample sizes we consider do not increase for the bivariate case, we expect to see a significant negative change in the performance of our estimator against the GARCH compared with the univariate MFY model, particularly in the case where $n = 500$.

5 Results

We considered two first stage estimators for VaR and ES: our nonparametric method and the ARMAX/Gaussian GARCH(1,1) model. For both methods, we only consider the L-moments procedure for the second stage. The considered estimators are based on stochastic models that are intentionally misspecified relative to our DGPs. The nonparametric model is assumed to depend only on Y_{t-1} and D_{t-1} (Markov property of order 1 with an exogenous variable). The GARCH model is misspecified because it assumes Gaussian innovations and because both g functions we consider are nonlinear functions of Y_{t-1} and D_{t-1}. In the local linear regression procedure for the first stage estimators, we use a Gaussian kernel with Ruppert's theoretical optimal bandwidth. A summary of the Monte Carlo simulation results for our estimator can be found in Appendix 1, Tables 2A and 2B. In Tables 3A and 3B, we present the results obtained from the MFY estimator. The motivation behind including the MFY estimator in the Monte Carlo is twofold: recreating the results of Martins-Filho and Yao (2006) and examining the effect of adding a regressor on the relative performance of the nonparametric estimator versus the GARCH method.

5.1 General Relative Performance

In general, the results for the MFY estimator are consistent with those of our multivariate estimator. One notable exception occurs, however, when we examine the case where $n = 500$. For both volatility structures, the MFY estimator outperforms GARCH in a similar number of experiments whether n is 1000 or 500. Our bivariate estimator, however, outperforms GARCH in significantly fewer experiments when $n = 500$. This result indicates that the curse of dimensionality exerts a significant effect on our estimator in small samples sizes. The improvement in performance between $n = 500$ and $n = 1000$ indicates that sufficient convergence of our estimator occurs for an n such that $500 < n < 1000$.

In almost all cases where $n = 1000$ and volatility is modeled by g_1, the nonparametric estimator for both VaR and ES outperforms GARCH on the basis of both MSE and bias. Outperformance is much less frequent in cases where $n = 500$. We notice for bias in particular that the nonparametric estimator outperforms GARCH in all but two cases. Given that our nonparametric estimator is inherently biased, this is an interesting result because it indicates that the nonlinearities present in the volatility function g_1 significantly hinder the performance of the GARCH estimator, which is asymptotically unbiased (Andrews (2009)). For the case where volatility is modeled by g_2, outperformance is witnessed far less frequently for both MSE and bias. An interesting pattern emerges in bias, however in that almost every experiment with $\gamma = 0.9$, the nonparametric estimator outperforms relative to GARCH. We do also see more frequent outperformance of the nonparametric estimator in these cases for RMSE, though the pattern isn't quite as stark as it is for bias. This result is unique to g_2 and also appears in our results for the MFY estimator. Our results indicate that though the nonlinearities of volatility are very important to the performance of our estimator, the γ coefficient also has a significant effect when the volatility is modeled by g_2. Additionally, we note that the nonparametric estimator is consistently less biased than the GARCH estimator for $\gamma = 0.9$, while it always underperforms for $\gamma = 0.3$.

Since both estimators are, by construction, misspecified to the actual distribution underlying the data, we expect performance to be related to the shape of the distribution. The results for both g_1 and g_2 support this, as our nonparametric estimator outperforms more frequently in experiments where $\lambda = -0.5$. Since our GARCH method is defined with normal innovations, it should perform poorly in estimating a skewed distribution, particularly one with heavy skewness such as the case where $\lambda = -0.5$. Our results support this hypothesis.

In both volatility structures, the nonparametric estimator outperformed more frequently for $\alpha = 0.99$ than for $\alpha = 0.95$ when estimating VaR. The opposite is true when estimating ES. This is because, by definition, ES is further out on the tail of the distribution than VaR, which makes ES more difficult to estimate. There is more variance in estimating ES than there is in estimating VaR. GARCH is less able to model VaR further out in the tails as well, which is why we see the nonparametric estimator outperform more frequently when $\alpha = 0.99$. Such an effect is muted for

ES. In fact, the nonparametric estimator outperforms GARCH in estimating ES in roughly half the experiments considered. We again attribute this to the curse of dimensionality since the results for the MFY estimator more consistently outperform GARCH for ES. Additionally, RMSE is generally much larger when estimating ES than when estimating VaR.

The number of observations used in the second stage, k, has no significant consistent impact on either the MSE or the bias of any of the estimators considered, ceteris paribus. This results supports the results of Martins-Filho and Yao (2006) and McNeil and Frey (2000).

5.2 Ceteris Paribus Relationships

The results of our Monte Carlo simulation allow us to make several *ceteris paribus* statements about the effect of our inputs on the performance of the bivariate nonparametric estimator.

- Sample size n: In general, as n increases, RMSE decreases for both volatility models and for both VaR and ES. For three of the four experiment pairings where $\lambda = 0.9$ and $\alpha = 0.99$, the relationship is reversed. The results for bias are mixed; there is no consistent discernible correlation between changing sample size and bias.

- Quantile α: Increasing the quantile from 0.95 to 0.99 increases both RMSE and bias in most experiments for both volatility models and for both VaR and ES. As Martins-Filho and Yao (2006) found, this result indicates that estimation of VaR and ES is more difficult for higher quantiles. For g_2, the relationship for RMSE is reversed for three out of four experiment pairings[4] where $\gamma = 0.3$ and $\lambda = -0.5$ for both VaR and ES. Such a reversal is not seen for bias.

- Lagged volatility weight γ: As expected, the RMSE for both VaR and ES in both volatility

[4]When referring to "experiment pairings," we mean groups of two experiments whose only difference is the parameter of interest. Grouping the experiments in this way is what allows our ceteris paribus analysis.

structures increases as γ goes from 0.3 to 0.9. This is true in all experiments. The same relationship holds for bias in all experiments in both volatility structures and for both VaR and for ES. In all experiment pairs where volatility is modeled by g_2 bias goes from negative to positive as γ goes from 0.3 to 0.9. The nonparametric estimator tends to underpredict VaR and ES when lagged volatility is weighted less and tends to overpredict VaR and ES when lagged volatility is weighted more heavily.

- Skewness parameter λ: The RMSE decreases significantly for all but one experiment as λ goes from 0 to -0.5. This relationship is likely explained by the fact that our DGP, when $\lambda \leq 0$, is skewed toward the positive quadrant. Therefore, in the second stage of our estimation procedure, when we select data larger than the k^{th} order statistic, we, by default, select data more representative of tail behavior when λ decreases. Contradictory to the results found by Martins-Filho and Yao (2006), there is a clear pattern that emerges in bias as λ decreases. In all but two experiment pairings for g_2, bias decreases along with λ. For g_1, a regular pattern emerges where for all experiments with $\gamma = 0.9$, decreasing λ decreases bias for both VaR and ES, while for all experiments where $\gamma = 0.3$, decreasing λ increases bias for both VaR and ES.

- Number of exceedances k: The impact of increasing k from 60 to 100 is unclear. For g_1, increasing k tends to increase RMSE and bias in a majority of cases, but the relationship is far from definitive. For g_2, the results are mixed enough that no obvious relationship emerges.

These results generally indicate that our bivariate nonparametric estimator outperforms GARCH in larger samples. By including a larger n, we notice significantly better performance in both volatility constructs, though the effect is more pronounced for $g_1(x)$.

6 Conclusion

In this paper we have proposed a modification of the method for estimating VaR and ES proposed by Martins-Filho and Yao (2006). Due to the popularity and widespread use of VaR and ES in the empirical and theoretical literature as well as in professional applications, a better understanding of market risk estimation is paramount to sound financial management. Our procedure extends the methodology used by Martins-Filho and Yao (2006) by generalizing it to the multivariate case, specifically by adding one exogenous variable. We used local linear regression techniques in stage one of our estimation procedure and L-moments and EVT in stage two to estimate the one-period forecasted VaR and ES. The Gaussian GARCH model is employed as an alternative first stage estimator for $\hat{m}$ and $\hat{\sigma^2}$ for comparative purposes. Our Monte Carlo simulation is based on a skewed Student-t distributed DGP that incorporates many of the empirically observed characteristics of financial returns series. The Monte Carlo simulation indicates that our estimation method outperforms the GARCH methodology, but does so much more consistently when $n = 1000$. We contend that what underperformance is present is due primarily to the curse of dimensionality. To our knowledge, this is the first evidence of the finite sample performance of VaR and ES estimators in multivariate conditional densities, particularly with consideration of exogenous variables. The results from our simulation indicate that nonlinearities in volatility dynamics exert a significant effect on estimates of risk. Concurrent with the findings of Martins-Filho and Yao (2006), this result indicates that accounting for the nonlinearities in volatility is more important than more comprehensive modeling of temporal dependence. More investigation is necessary, however, to determine the performance of our estimator in a greater variety of parameter configurations.

Areas for Further Research: There remain several avenues for further investigation into the properties of our estimators. The extent of our Monte Carlo simulation was truncated due to limitations on computing power and time. Future researchers may find examining additional cases in the simulation, particularly those which consider larger sample sizes, to yield interesting results. Additionally, consideration of a larger variety of GARCH variants would reveal a more complete picture of the relative performance of our estimators. Researchers more concerned with the local regression procedure may also be interested in exploring the performance of a local polynomial estimators, assuming continuity assumptions are relaxed. With greater computing power and more time, it may also be interesting to examine the effects of adding more lags and more exogenous variables to the DGP. Researchers examining this question, however, would need to consider extremely large sample sizes. That said, a comprehensive backtesting evaluation of the estimator on actual historical financial time series would provide a glimpse into the real-world performance of our estimator compared to the methods currently employed by practitioners. One useful modification of this model would be to redefine the estimator in terms of an additive model. Doing so would mitigate the impact of the curse of dimensionality and allow the estimator to converge at nearly the rate of the GARCH model (Andersen et al. (2009)). In fact, the best possible rate of convergence for estimates of σ_t^2 is equal to that of the univariate nonparametric regression (Stone (1985)). The additive version, is, however, more restrictive on the functional form of the estimator. Finally, like the method set forth by Martins-Filho and Yao (2006), the asymptotic characteristics of these estimators are also yet unknown and could prove interesting to the theoretical researcher.

A Appendix A - Tables and Graphs

Figure 1: Conditional Volatility based on
$g_1(x)$ **and** $g_2(x)$ **where** $x = X(Y_{t-1}, d_{t-1})$

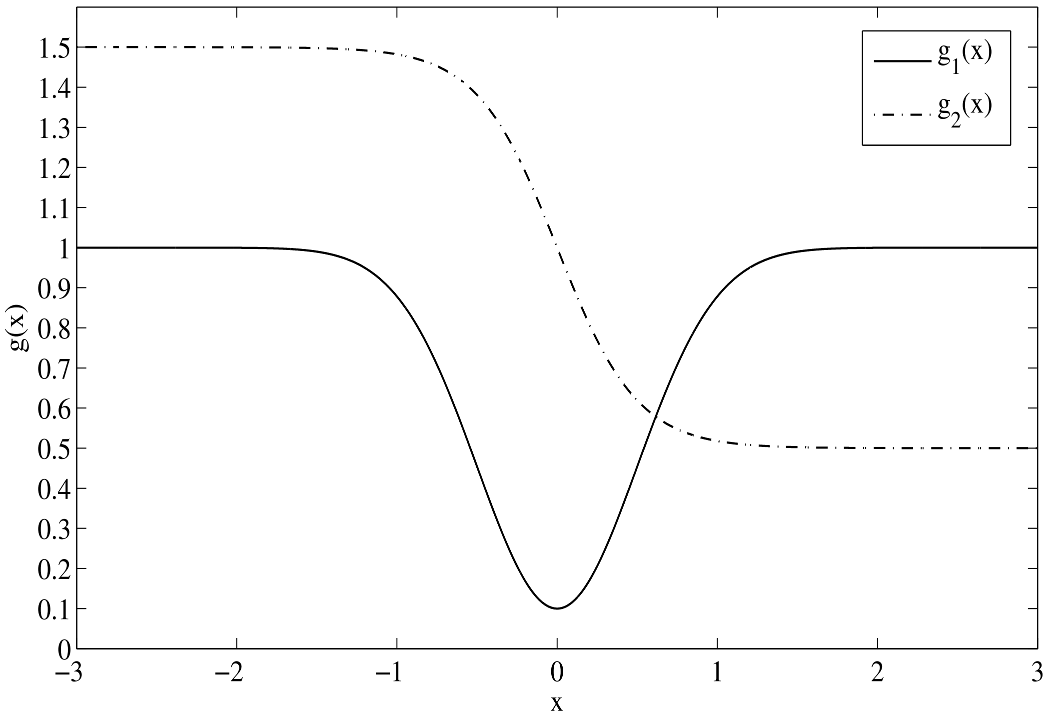

Table 1: Numbering of Experiments
Volatility based on g_1, g_2

Exp	λ	n	γ	α	k
1	0	1000	0.3	0.95	60
2	-0.5	1000	0.3	0.95	60
3	0	1000	0.9	0.95	60
4	-0.5	1000	0.9	0.95	60
5	0	1000	0.3	0.99	60
6	-0.5	1000	0.3	0.99	60
7	0	1000	0.9	0.99	60
8	-0.5	1000	0.9	0.99	60
9	0	1000	0.3	0.95	100
10	-0.5	1000	0.3	0.95	100
11	0	1000	0.9	0.95	100
12	-0.5	1000	0.9	0.95	100
13	0	1000	0.3	0.99	100
14	-0.5	1000	0.3	0.99	100
15	0	1000	0.9	0.99	100
16	-0.5	1000	0.9	0.99	100
17	0	500	0.3	0.95	60
18	-0.5	500	0.3	0.95	60
19	0	500	0.9	0.95	60
20	-0.5	500	0.9	0.95	60
21	0	500	0.3	0.99	60
22	-0.5	500	0.3	0.99	60
23	0	500	0.9	0.99	60
24	-0.5	500	0.9	0.99	60
25	0	500	0.3	0.95	100
26	-0.5	500	0.3	0.95	100
27	0	500	0.9	0.95	100
28	-0.5	500	0.9	0.95	100
29	0	500	0.3	0.99	100
30	-0.5	500	0.3	0.99	100
31	0	500	0.9	0.99	100
32	-0.5	500	0.9	0.99	100

Table 2A: Root MSE and Bias
Volatility based on g_1
NP - Nonparametric estimator, G - GARCH Estimator

Exp	VaR						ES					
	RMSE			Bias			RMSE			Bias		
	Ratio	NP	G	Ratio	NP	G	Ratio	NP	G	Ratio	NP	G
1	0.766	0.193	0.252	0.288	0.019	0.066	0.749	0.263	0.351	0.083	0.009	0.109
2	0.707	0.217	0.307	0.299	0.023	0.077	0.663	0.248	0.374	0.342	0.039	0.114
3	0.974	1.912	1.963	0.953	1.836	1.927	0.952	2.552	2.681	0.933	2.451	2.627
4	0.994	1.454	1.463	0.967	1.393	1.441	0.992	1.768	1.783	0.969	1.702	1.756
5	0.671	0.302	0.450	0.133	0.019	0.143	0.767	0.488	0.636	0.074	0.015	0.204
6	0.754	0.214	0.284	0.462	0.048	0.104	0.963	0.337	0.350	0.558	0.072	0.129
7	0.962	2.974	3.093	0.928	2.820	3.038	0.946	3.641	3.848	0.913	3.401	3.726
8	0.988	2.009	2.034	0.962	1.927	2.003	1.009	2.402	2.381	0.979	2.276	2.324
9	0.720	0.175	0.243	0.346	0.018	0.052	0.671	0.228	0.340	0.102	0.010	0.098
10	0.640	0.126	0.197	0.517	0.030	0.058	0.656	0.162	0.247	0.505	0.047	0.093
11	0.995	1.944	1.953	0.963	1.850	1.922	0.968	2.580	2.666	0.937	2.456	2.621
12	1.023	1.563	1.528	0.951	1.412	1.485	1.011	1.884	1.864	0.956	1.734	1.814
13	0.720	0.357	0.496	0.069	0.009	0.130	0.937	0.640	0.683	0.230	-0.035	0.152
14	0.766	0.219	0.286	0.618	0.068	0.110	0.921	0.313	0.340	0.645	0.069	0.107
15	0.964	3.008	3.120	0.951	2.918	3.068	0.936	3.645	3.896	0.923	3.476	3.765
16	1.004	2.048	2.040	0.974	1.954	2.007	1.021	2.380	2.330	0.985	2.234	2.267
17	1.084	0.297	0.274	0.123	0.007	0.057	0.972	0.381	0.392	0.368	-0.032	0.087
18	1.177	0.272	0.231	0.185	0.012	0.065	1.076	0.310	0.288	0.320	0.032	0.100
19	0.987	1.988	2.015	0.913	1.790	1.960	0.940	2.608	2.773	0.874	2.354	2.692
20	0.990	1.476	1.491	0.954	1.392	1.459	1.007	1.821	1.808	0.972	1.716	1.766
21	1.020	0.502	0.492	0.563	-0.071	0.126	1.043	0.745	0.714	1.209	-0.162	0.134
22	0.824	0.258	0.313	0.292	0.035	0.120	0.937	0.357	0.381	0.183	0.022	0.120
23	0.922	2.805	3.041	0.881	2.594	2.943	0.880	3.336	3.789	0.837	2.963	3.538
24	1.024	2.108	2.059	0.977	1.964	2.010	1.061	2.501	2.358	0.993	2.247	2.262
25	1.084	0.296	0.273	0.210	0.013	0.062	0.995	0.385	0.387	0.554	-0.046	0.083
26	1.013	0.232	0.229	0.359	0.028	0.078	1.044	0.283	0.271	0.300	0.027	0.090
27	1.012	2.015	1.991	0.963	1.872	1.944	0.978	2.623	2.683	0.931	2.431	2.610
28	1.008	1.582	1.569	0.940	1.449	1.541	1.032	1.913	1.854	0.959	1.740	1.815
29	0.946	0.459	0.485	0.426	-0.063	0.148	1.001	0.681	0.680	2.404	-0.238	0.099
30	1.016	0.326	0.321	0.364	0.043	0.118	1.263	0.485	0.384	0.185	-0.012	0.065
31	0.936	2.930	3.131	0.861	2.614	3.035	0.897	3.314	3.696	0.815	2.829	3.473
32	0.998	2.030	2.034	0.958	1.894	1.978	1.010	2.182	2.160	0.959	1.974	2.058

Table 2B: Root MSE and Bias
Volatility based on g_2
NP - Nonparametric estimator, G - GARCH Estimator

Exp	VaR						ES					
	RMSE			Bias			RMSE			Bias		
	Ratio	NP	G	Ratio	NP	G	Ratio	NP	G	Ratio	NP	G
1	1.530	0.306	0.200	1.657	-0.111	-0.067	1.452	0.379	0.261	3.298	-0.155	-0.047
2	1.281	0.260	0.203	1.484	-0.092	-0.062	1.125	0.270	0.240	3.536	-0.099	-0.028
3	0.983	1.191	1.212	0.949	1.106	1.166	0.969	1.611	1.662	0.934	1.490	1.596
4	1.004	0.996	0.992	0.968	0.924	0.955	1.008	1.224	1.214	0.977	1.145	1.172
5	1.108	0.390	0.352	2.813	-0.211	-0.075	1.099	0.500	0.455	5.520	-0.276	-0.050
6	0.704	0.197	0.280	6.714	-0.094	0.014	0.705	0.249	0.353	1.712	-0.101	0.059
7	0.974	1.875	1.925	0.927	1.710	1.844	0.963	2.281	2.368	0.911	2.031	2.229
8	0.993	1.332	1.342	0.974	1.276	1.310	1.037	1.609	1.552	0.999	1.500	1.501
9	1.430	0.329	0.230	1.487	-0.116	-0.078	1.327	0.398	0.300	2.842	-0.162	-0.057
10	0.941	0.144	0.153	1.950	-0.078	-0.040	0.918	0.168	0.183	83.000	-0.083	-0.001
11	0.998	1.217	1.219	0.909	1.061	1.167	0.959	1.610	1.679	0.891	1.434	1.609
12	1.007	1.020	1.013	0.924	0.897	0.971	0.994	1.235	1.242	0.933	1.109	1.189
13	1.155	0.409	0.354	3.322	-0.196	-0.059	1.078	0.497	0.461	8.700	-0.261	-0.030
14	0.812	0.177	0.218	3.895	-0.074	0.019	0.815	0.224	0.275	1.339	-0.083	0.062
15	0.937	1.831	1.955	0.904	1.697	1.877	0.907	2.181	2.405	0.877	1.969	2.246
16	1.012	1.401	1.384	0.993	1.330	1.340	1.048	1.638	1.563	1.012	1.513	1.495
17	1.410	0.323	0.229	1.795	-0.149	-0.083	1.385	0.417	0.301	3.129	-0.219	-0.070
18	1.829	0.353	0.193	2.892	-0.107	-0.037	1.622	0.378	0.233	38.000	-0.114	0.003
19	1.014	1.257	1.240	0.921	1.075	1.167	0.965	1.637	1.696	0.883	1.408	1.595
20	0.977	1.011	1.035	0.934	0.922	0.987	0.987	1.252	1.269	0.941	1.141	1.212
21	1.381	0.526	0.381	5.064	-0.238	-0.047	1.235	0.642	0.520	8.525	-0.341	-0.040
22	0.992	0.256	0.258	6.667	-0.100	0.015	1.021	0.340	0.333	2.170	-0.115	0.053
23	0.945	1.834	1.941	0.859	1.570	1.827	0.906	2.216	2.445	0.818	1.800	2.201
24	1.020	1.466	1.437	0.899	1.237	1.376	1.044	1.681	1.610	0.907	1.369	1.509
25	1.147	0.266	0.232	2.052	-0.119	-0.058	1.088	0.347	0.319	3.439	-0.196	-0.057
26	1.726	0.302	0.175	3.241	-0.094	-0.029	1.561	0.331	0.212	20.800	-0.104	0.005
27	0.990	1.271	1.284	0.895	1.085	1.212	0.944	1.593	1.687	0.856	1.350	1.577
28	1.012	1.080	1.067	0.915	0.932	1.019	1.002	1.265	1.262	0.919	1.107	1.205
29	1.337	0.544	0.407	3.687	-0.247	-0.067	1.307	0.707	0.541	4.063	-0.386	-0.095
30	1.156	0.267	0.231	4.889	-0.088	0.018	1.117	0.324	0.290	5.440	-0.136	0.025
31	0.961	1.889	1.966	0.881	1.629	1.850	0.930	2.171	2.335	0.832	1.746	2.098
32	1.052	1.418	1.348	0.996	1.282	1.287	1.101	1.559	1.416	1.017	1.334	1.312

Table 3A: Root MSE and Bias

Volatility based on g_1

MFY - MFY Nonparametric estimator, G - GARCH Estimator

Exp	VaR						ES					
	RMSE			Bias			RMSE			Bias		
	Ratio	MFY	G	Ratio	MFY	G	Ratio	MFY	G	Ratio	MFY	G
1	0.551	0.158	0.287	0.522	0.036	0.069	0.559	0.223	0.399	0.372	0.045	0.121
2	0.417	0.130	0.312	0.439	0.036	0.082	0.426	0.163	0.383	0.391	0.050	0.128
3	0.982	1.934	1.969	0.984	1.902	1.933	0.970	2.614	2.694	0.974	2.569	2.638
4	0.990	1.441	1.455	0.987	1.415	1.434	0.988	1.756	1.778	0.985	1.726	1.752
5	0.537	0.269	0.501	0.393	0.055	0.140	0.597	0.412	0.690	0.332	0.069	0.208
6	0.503	0.163	0.324	0.440	0.055	0.125	0.578	0.227	0.393	0.410	0.064	0.156
7	0.981	3.039	3.099	0.971	2.955	3.043	0.972	3.754	3.864	0.961	3.593	3.739
8	0.986	1.990	2.018	0.982	1.949	1.985	0.982	2.332	2.375	0.978	2.266	2.316
9	0.574	0.155	0.270	0.698	0.037	0.053	0.534	0.204	0.382	0.421	0.045	0.107
10	0.452	0.100	0.221	0.544	0.031	0.057	0.462	0.129	0.279	0.460	0.046	0.100
11	1.004	1.957	1.950	0.998	1.917	1.921	0.991	2.645	2.670	0.986	2.588	2.625
12	0.967	1.469	1.519	0.968	1.431	1.478	0.965	1.792	1.857	0.968	1.750	1.808
13	0.565	0.286	0.506	0.439	0.058	0.132	0.632	0.428	0.677	0.116	0.018	0.155
14	0.532	0.173	0.325	0.617	0.074	0.120	0.605	0.233	0.385	0.492	0.062	0.126
15	0.982	3.087	3.144	0.980	3.026	3.089	0.966	3.784	3.918	0.965	3.656	3.788
16	0.996	2.018	2.027	0.994	1.983	1.995	1.003	2.321	2.314	0.998	2.249	2.254
17	0.743	0.231	0.311	0.383	0.023	0.060	0.713	0.316	0.443	0.228	0.023	0.101
18	0.580	0.145	0.250	0.427	0.032	0.075	0.615	0.193	0.314	0.415	0.049	0.118
19	0.995	2.026	2.036	0.965	1.914	1.983	0.972	2.721	2.800	0.945	2.573	2.722
20	1.005	1.478	1.470	0.998	1.440	1.443	1.003	1.793	1.788	0.997	1.744	1.749
21	0.735	0.378	0.514	0.287	0.035	0.122	0.789	0.587	0.744	0.162	-0.024	0.148
22	0.702	0.236	0.336	0.482	0.066	0.137	0.781	0.317	0.406	0.382	0.055	0.144
23	0.980	3.007	3.067	0.977	2.904	2.972	0.964	3.693	3.830	0.958	3.432	3.582
24	0.989	2.040	2.063	0.985	1.979	2.009	0.984	2.338	2.377	0.977	2.221	2.274
25	0.746	0.229	0.307	0.703	0.052	0.074	0.704	0.305	0.433	0.300	0.033	0.110
26	0.916	0.228	0.249	0.367	0.033	0.090	0.880	0.264	0.300	0.245	0.027	0.110
27	1.014	2.031	2.003	1.006	1.969	1.958	1.001	2.696	2.694	0.992	2.603	2.623
28	0.996	1.539	1.545	0.984	1.496	1.520	0.997	1.827	1.832	0.982	1.763	1.796
29	0.682	0.369	0.541	0.249	0.043	0.173	0.751	0.562	0.748	0.528	-0.075	0.142
30	0.683	0.235	0.344	0.376	0.053	0.141	0.894	0.371	0.415	0.065	-0.007	0.107
31	0.971	3.058	3.150	0.950	2.903	3.055	0.956	3.570	3.736	0.933	3.277	3.511
32	0.985	1.982	2.013	0.970	1.900	1.959	0.983	2.106	2.142	0.967	1.971	2.039

Table 3B: Root MSE and Bias

Volatility based on g_2

MFY - MFY Nonparametric estimator, G - GARCH Estimator

Exp	VaR						ES					
	RMSE			Bias			RMSE			Bias		
	Ratio	MFY	G	Ratio	MFY	G	Ratio	MFY	G	Ratio	MFY	G
1	0.845	0.169	0.200	1.478	-0.102	-0.069	0.845	0.223	0.264	1.523	-0.131	-0.086
2	0.589	0.172	0.292	2.074	-0.056	-0.027	0.526	0.195	0.371	2.280	-0.057	-0.025
3	1.014	1.556	1.534	0.988	1.486	1.504	0.996	2.093	2.101	0.986	2.025	2.053
4	0.990	1.246	1.258	0.991	1.224	1.235	0.995	1.525	1.532	0.995	1.497	1.504
5	0.922	0.366	0.397	2.026	-0.154	-0.076	0.880	0.461	0.524	1.918	-0.188	-0.098
6	0.493	0.176	0.357	3.188	-0.051	-0.016	0.496	0.210	0.423	2.500	-0.050	-0.020
7	0.991	2.431	2.453	0.975	2.334	2.395	0.987	2.994	3.033	0.970	2.829	2.917
8	0.992	1.684	1.697	0.985	1.650	1.675	1.002	1.973	1.969	0.991	1.910	1.928
9	1.221	0.276	0.226	1.925	-0.102	-0.053	1.059	0.321	0.303	1.941	-0.132	-0.068
10	0.622	0.102	0.164	1.104	-0.053	-0.048	0.622	0.122	0.196	0.982	-0.054	-0.055
11	0.988	1.510	1.529	0.982	1.468	1.495	0.980	2.061	2.102	0.973	1.999	2.055
12	0.983	1.255	1.277	0.981	1.225	1.249	0.979	1.522	1.554	0.980	1.487	1.518
13	0.646	0.263	0.407	1.577	-0.123	-0.078	0.690	0.365	0.529	1.518	-0.173	-0.114
14	0.609	0.123	0.202	1.083	-0.052	-0.048	0.690	0.167	0.242	0.954	-0.062	-0.065
15	0.985	2.417	2.454	0.966	2.324	2.407	0.973	2.937	3.020	0.954	2.773	2.906
16	1.005	1.724	1.716	0.999	1.690	1.691	1.006	1.947	1.936	0.997	1.884	1.889
17	1.120	0.224	0.200	1.629	-0.114	-0.070	1.133	0.299	0.264	1.625	-0.156	-0.096
18	0.847	0.177	0.209	1.641	-0.064	-0.039	0.804	0.201	0.250	1.591	-0.070	-0.044
19	1.001	1.566	1.564	0.976	1.475	1.512	0.985	2.107	2.139	0.962	1.985	2.064
20	0.994	1.277	1.285	0.985	1.230	1.249	0.994	1.559	1.568	0.982	1.499	1.526
21	0.804	0.401	0.499	2.387	-0.179	-0.075	0.874	0.535	0.612	1.881	-0.252	-0.134
22	1.036	0.348	0.336	1.920	-0.048	-0.025	1.024	0.388	0.379	1.400	-0.063	-0.045
23	0.976	2.405	2.464	0.954	2.263	2.373	0.965	2.965	3.074	0.942	2.694	2.859
24	0.988	1.757	1.778	0.975	1.686	1.729	0.989	1.963	1.985	0.973	1.846	1.898
25	1.203	0.273	0.227	1.745	-0.089	-0.051	1.057	0.332	0.314	1.728	-0.140	-0.081
26	0.649	0.126	0.194	2.222	-0.060	-0.027	0.675	0.156	0.231	1.868	-0.071	-0.038
27	0.991	1.587	1.602	0.974	1.510	1.551	0.979	2.065	2.110	0.961	1.947	2.027
28	0.992	1.314	1.325	0.985	1.275	1.294	0.994	1.558	1.567	0.987	1.506	1.526
29	1.210	0.478	0.395	1.714	-0.132	-0.077	1.144	0.602	0.526	1.461	-0.241	-0.165
30	0.652	0.176	0.270	1.683	-0.069	-0.041	0.753	0.238	0.316	1.253	-0.119	-0.095
31	0.993	2.462	2.480	0.965	2.295	2.379	0.980	2.877	2.937	0.948	2.566	2.708
32	1.008	1.706	1.693	1.002	1.653	1.649	1.017	1.803	1.773	1.008	1.704	1.691

B References

Acerbi, C., Tasche, D., 2002. Expected Shortfall: A natural coherent alternative to value at risk. Economic Notes 31 (2), 379–388.

Ait-Sahalia, Y., Brandt, M., 2001. Variable selection for portfolio choice. The Journal of Finance 56 (4), 1297–1351.

Andersen, T., Benzoni, L., 2010. Do bonds span volatility risk in the US Treasury market? A specification test for affine term structure models. The Journal of Finance 65 (2), 603–653.

Andersen, T., Bollerslev, T., 1997. Heterogeneous information arrivals and return volatility dynamics: Uncovering the long-run in high frequency returns. Journal of finance 52 (3), 975–1005.

Andersen, T., Bollerslev, T., 1998. Deutsche mark/dollar volatility: Intraday activity patterns, macroeconomic announcements, and longer run dependencies. the Journal of Finance 53 (1), 219–265.

Andersen, T., Bollerslev, T., Diebold, F., Ebens, H., 2001. The distribution of realized stock return volatility. Journal of Financial Economics 61 (1), 43–76.

Andersen, T., Davis, R., Kreiß, J., Apr. 2009. Handbook of financial time series. Vol. 78. Springer Verlag, Berlin.

Andreou, E., Pittis, N., Spanos, A., 2001. On modeling speculative prices: The empirical literature. Journal of Economic Surveys 15 (2), 187–220.

Andrews, B., 2009. Rank based estimation for GARCH processes. Tech. rep., Discussion Paper, Colorado State University.

Areal, N., Taylor, S., 2002. The realized volatility of FTSE-100 futures prices. Journal of Futures Markets 22 (7), 627–648.

Artzner, P., Delbaen, F., Eber, J., Heath, D., 1999. Coherent measures of risk. Mathematical finance 9 (3), 203–228.

Billingsley, P., 1961. Statistical inference for markov processes. University of Chicago Press, Chicago.

Black, F., 1976. Studies of stock price volatility changes. In: Proceedings of the 1976 meetings of the business and economic statistics section, American Statistical Association. Vol. 177. pp. 177–181.

Bodie, Z., Kane, A., Marcus, A., 2009. Investments, 8th Edition. McGraw-Hill, New York.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31 (3), 307–327.

Bollerslev, T., 2007. Glossary to ARCH (GARCH). The Journal Of The Bertrand Russell Archives.

Bollerslev, T., Chou, R. Y., Kroner, K. F., May 1992. ARCH modeling in finance: A review of the theory and empirical evidence. Journal of Econometrics 52 (1-2), 5–59.

Breidt, F., Crato, N., De Lima, P., 1998. The detection and estimation of long memory in stochastic volatility. Journal of Econometrics 83 (1-2), 325–348.

Campbell, J., Lo, A., MacKinlay, A., 1997. The econometrics of financial markets. Vol. 1. Princeton University Press, Princeton, NJ.

Carroll, R., Hardle, W., Mammen, E., 2002. Estimation in an additive model when the components are linked parametrically. Econometric Theory 18 (04), 886–912.

Chan, N., Wei, C., 1988. Limiting distributions of least squares estimates of unstable autoregressive processes. The Annals of Statistics 16 (1), 367–401.

Chen, Y., 2001. Testing conditional symmetry with an application to stock returns. Institute for Social Science and Philosophy Academia.

Cheung, Y., 1993. Long memory in foreign-exchange rates. Journal of Business & Economic Statistics 11 (1), 93–101.

Chow, K., Pan, M., Sakano, R., 1996. On the long-term or short-term dependence in stock prices: Evidence from international stock markets. Review of Quantitative Finance and Accounting 6 (2), 181–194.

Christie, A., 1982. The stochastic behavior of common stock variances:: Value, leverage and interest rate effects. Journal of Financial Economics 10 (4), 407–432.

Dacunha-Castelle, D., Florens-Zmirou, D., 1986. Estimation of the coefficients of a diffusion from discrete observations. Stochastics An International Journal of Probability and Stochastic Processes 19 (4), 263–284.

Davidian, M., Carroll, R., 1987. Variance function estimation. Journal of the American Statistical Association 82 (400), 1079–1091.

Deo, R., Hurvich, C., Lu, Y., 2006. Forecasting realized volatility using a long-memory stochastic volatility model: Estimation, prediction and seasonal adjustment. Journal of Econometrics 131 (1-2), 29–58.

Diebold, F. X., Schuermann, T., Stroughair, J. D., 1993. Pitfalls and opportunities in the use of Extreme Value Theory in risk management. In: Refenes, A.-P. N., Moody, J., Burgess, A. (Eds.), Advances in Computational Finance. Vol. 1. Kluwer Academic Publishers, Amsterdam, pp. 3–12.

Diebolt, J., Guégan, D., 1993. Tail behaviour of the stationary density of general non-linear autoregressive processes of order 1. Journal of Applied Probability 30 (2), 315–329.

Ding, Z., Granger, C., Engle, R., 1993. A long memory property of stock market returns and a new model. Journal of Empirical Finance 1 (1), 83–106.

Engle, R., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica: Journal of the Econometric Society, 987–1007.

Engle, R., Lee, G., 1999. A long-run and short-run component model of stock return volatility. Oxford University Press, Oxford.

Engle, R., Mustafa, C., 1992. Implied ARCH models from options prices. Journal of Econometrics 52 (1-2), 289–311.

Engle, R., Ng, V., 1993. Measuring and testing the impact of news on volatility. Journal of finance, 1749–1778.

Engle, R., Patton, A., Feb. 2001. What good is a volatility model? Quantitative Finance 1 (2), 237–245.

Fama, E., 1965. The behavior of stock-market prices. Journal of Business 38 (1), 34–105.

Fan, J., Dec. 1992. Design-adaptive nonparametric regression. Journal of the American Statistical Association 87 (420), 998.

Fan, J., Yao, Q., 1998. Efficient estimation of conditional variance functions in stochastic regression. Biometrika 85 (3), 645–660.

Fisher, R. A., Tippett, L. H. C., 1928. Limiting forms of the frequency distribution of the largest or smallest member of a sample. In: Proceedings of the Cambridge Philosophical Society. Vol. 24. pp. 180–190.

Gallant, A., Rossi, P., Tauchen, G., 1992. Stock prices and volume. Review of Financial studies 5 (2), 199.

Gallant, A., Tauchen, G., 1989. Seminonparametric estimation of conditionally constrained heterogeneous processes: Asset pricing applications. Econometrica: Journal of the Econometric Society 57 (5), 1091–1120.

Gnedenko, B. V., 1943. Sur la distribution limite du terme maximum d'une serie aleatoire. Annals of Mathematics 44 (3), 423–453.

Goetzmann, W., 1993. Patterns in three centuries of stock market prices. Journal of Business 66 (2), 249–270.

Gouriéroux, C., 1997. ARCH models and financial applications. Springer Verlag.

Grau-Carles, P., 2005. Tests of long memory: a bootstrap approach. Computational Economics 25 (1), 103–113.

Hafner, C., 1998. Nonlinear time series analysis with applications to foreign exchange rate volatility. Physica-Verlag Heidelberg, Heidelberg.

Hall, P., Heyde, C., 1980. Martingale limit theory and its application. Academic Press, New York.

Hansen, B., 1994. Autoregressive conditional density estimation. International Economic Review, 705–730.

Hardle, W., Tsybakov, A., 1997. Local polynomial estimators of the volatility function in nonparametric autoregression. Journal of Econometrics 81 (1), 223–242.

Holan, S. H., Lund, R., Davis, G., 2010. The ARMA alphabet soup: A tour of ARMA model variants. Statistics Surveys 4 (November 2009), 232–274.

Hosking, J., 1989. Some theoretical results concerning L-moments. IBM TJ Watson Research Center.

Hosking, J., 1990. L-moments: analysis and estimation of distributions using linear combinations of order statistics. Journal of the Royal Statistical Society. Series B (Methodological) 52 (1), 105–124.

Hosking, J., Wallis, J., 1987. Parameter and quantile estimation for the generalized pareto distribution. Technometrics 29 (3), 339–349.

Hosking, J., Wallis, J., 1997. Regional frequency analysis: An approach based on L-moments. Cambridge Univ Pr.

Hosking, J., Wallis, J., Wood, E., 1985. Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. Technometrics 27 (3), 251–261.

Huang, B., Yang, C., 1999. An examination of long-term memory using the intraday stock returns. Tech. rep., Technical Report 9903, Clarion University of Pennsylvania, Clarion.

IBM Corporation, 2003. The L-moments page.
URL `http://www.research.ibm.com/people/h/hosking/lmoments.html`

Jorion, P., 2001. Value at risk. Vol. 2. McGraw-Hill New York.

JPMorgan, 1996. RiskMetrics - Technical document. Tech. rep., JP Morgan, New York.

Kroner, K. F., Ng, V. K., 1998. Modeling asymmetric comovements of asset returns. Review of Financial Studies 11 (4), 817–844.

Kupiec, P., 1989. Initial margin requirements and stock returns volatility: Another look. Journal of Financial Services Research 3 (2), 287–301.

Leontaritis, I., Billings, S., 1985a. Input-output parametric models for non-linear systems part I: deterministic non-linear systems. International Journal of Control 41 (2), 303–328.

Leontaritis, I., Billings, S., 1985b. Input-output parametric models for non-linear systems part II: stochastic non-linear systems. International Journal of Control 41 (2), 329–344.

Li, Q., Racine, J., 2007. Nonparametric econometrics: Theory and practice. Princeton University Press, Princeton, NJ.

Limam, I., 2003. Is long memory a property of thin stock markets? International evidence using Arab countries. Review of Middle East Economics and Finance 1 (3), 251–266.

Lo, A., 1991. Long-term memory in stock market prices. Econometrica: Journal of the Econometric Society 59 (5), 1279–1313.

Lobato, I., Savin, N., 1998. Real and spurious long-memory properties of stock-market data. Journal of Business & Economic Statistics 16 (3), 261–268.

Mandelbrot, B., 1963. The variation of certain speculative prices. Journal of Business 36 (4), 394–419.

Martens, M., 2002. Measuring and forecasting S&P 500 index-futures volatility using high-frequency data. Journal of Futures Markets 22 (6), 497–518.

Martins-Filho, C., Saraiva, P., 2011. On asymptotic normality of the local polynomial regression estimator with stochastic bandwidths. Communications in Statistics - Theory and Methods, forthcoming.

Martins-Filho, C., Yao, F., 2006. Estimation of value-at-risk and expected shortfall based on nonlinear models of return dynamics and extreme value theory. Studies in Nonlinear Dynamics & Econometrics 10 (2), 1304.

Masry, E., 1995. Nonparametric estimation and identification of nonlinear ARCH time series. Econometric Theory 11, 258–289.

McFadden, D., 1978. Modeling the choice of residential location. Transportation Research Record (673).

McNeil, A., Frey, R., Nov. 2000. Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. Journal of Empirical Finance 7 (3-4), 271–300.

Milhoj, A., 1985. The moment structure of ARCH processes. Scandinavian Journal of Statistics 12 (4), 281–292.

Nawrocki, D., 1993. R/S analysis and long term dependence in stock market indices. Managerial Finance 21 (7), 78–91.

Nelson, D., 1990. Stationarity and persistence in the GARCH(1,1) model. Econometric Theory 6 (03), 318–334.

Oh, G., Um, C., Kim, S., 2006. Long-term memory and volatility clustering in daily and high-frequency price changes. Arxiv preprint physics/0601174.

Oja, H., 1981. On location, scale, skewness and kurtosis of univariate distributions. Scandinavian Journal of Statistics 8 (3), 154–168.

Pagan, A., Schwert, G. W., 1990. Alternative models for conditional volatility. Journal of Econometrics 45, 267–290.

Patton, A., 2004. On the importance of skewness and asymmetric dependence in stock returns for asset allocation. forthcoming, Journal of Financial Econometrics.

Peng, H., Ozaki, T., Toyoda, Y., Oda, K., 2001. Modeling and control of systems with signal dependent nonlinear dynamics. Feedback 3.

Phillips, P., Yu, J., 2009. A two-stage realized volatility approach to estimation of diffusion processes with discrete data. Journal of Econometrics 150 (2), 139–150.

Pickands, J., 1975. Statistical inference using extreme order statistics. The Annals of Statistics, 119–131.

Posedel, P., 2006. Analysis of the exchange rate and pricing foreign currency options on the Croatian market: The NGARCH model as an alternative to the Black-Scholes model. Financial Theory and Practice 30 (4), 347–368.

Poterba, J., Summers, L., 1986. The persistence of volatility and stock market fluctuations. The American Economic Review 76 (5), 1142–1151.

Royston, P., 1992. Which measures of skewness and kurtosis are best? Statistics in medicine 11 (3), 333–343.

Ruppert, D., Sheather, S., Wand, M., Dec. 1995. An effective bandwidth selector for local least squares regression. Journal of the American Statistical Association 90 (432), 1257–1270.

Sheth, B. A., Kim, D., 2003. The alphabet-soup of ARCH models. Image (Rochester, N.Y.).

Stone, C., 1980. Optimal rates of convergence for nonparametric estimators. The Annals of Statistics.

Stone, C., 1985. Additive regression and other nonparametric models. The Annals of Statistics.

Tauchen, G., 2001. Notes on financial econometrics. Journal of Econometrics 100 (1), 57–64.

Terasvirta, T., 2008. An introduction to univariate GARCH models. In: Anderson, T. G., Davis, R. A., Kreiss, J.-P., Mikosch, T. (Eds.), Handbook of Financial Time Series. Springer, New York, pp. 17–42.

Teräsvirta, T., Zhao, Z., 2006. Stylized facts of return series, robust estimates, and three popular models of volatility. Unpublished paper, Stockholm School of Economics.

Wald, A., 1949. Note on the consistency of the maximum likelihood estimate. The Annals of Mathematical Statistics 20 (4), 595–601.

Wang, Q., 1997. Direct sample estimators of L-moments. Water Resources Research 32 (12), 3617–3619.

Zakoian, J., 1994. Threshold heteroskedastic models. Journal of Economic Dynamics and control 18 (5), 931–955.

Zumbach, G., 2004. Volatility processes and volatility forecast with long memory. Quantitative Finance 4 (1), 70–86.